

Quantum machine learning for data scientists

Dawid Kopczyk

Quantee Limited, Manchester, United Kingdom

This text aims to present and explain quantum machine learning algorithms to a data scientist in an accessible and consistent way. The algorithms and equations presented are not written in rigorous mathematical fashion, instead, the pressure is put on examples and step by step explanation of difficult topics. This contribution gives an overview of selected quantum machine learning algorithms, however there is also a method of scores extraction for quantum PCA algorithm proposed as well as a new cost function in feed-forward quantum neural networks is introduced. The text is divided into four parts: the first part explains the basic quantum theory, then quantum computation and quantum computer architecture are explained in section two. The third part presents quantum algorithms which will be used as subroutines in quantum machine learning algorithms. Finally, the fourth section describes quantum machine learning algorithms with the use of knowledge accumulated in previous parts.

Contents

I. Introduction	2
II. Basic quantum theory	3
A. Quantum states	3
B. Quantum observables	5
C. Measurement	7
D. Assembling quantum states	8
E. Density matrix	9
III. Quantum computation	11
A. Qubit	11
B. Quantum gates	12
C. Quantum parallelism	14
IV. Quantum algorithms	15
A. Grover's search algorithm	16
1. Preliminaries	16
2. Example	16
3. Algorithm	19
4. Summary	19
B. Quantum minimization algorithm	19
1. Preliminaries	19
2. Example	20
3. Algorithm	22
4. Summary	22
C. Quantum Fourier transform algorithm	23
1. Preliminaries	23
2. Example	25
3. Algorithm	26
4. Summary	27
D. Quantum phase estimation algorithm	27
1. Preliminaries	27
2. Example	28
3. Algorithm	30
4. Summary	31
V. Quantum machine learning algorithms	31
A. Quantum k -means algorithm	32

1. Preliminaries	32
2. Quantum subroutine: <i>SwapTest</i>	32
3. Quantum subroutine: <i>DistCalc</i>	34
4. Algorithm	35
5. Summary	35
B. Quantum k -medians algorithm	36
1. Preliminaries	36
2. Quantum subroutine: <i>MedianCalc</i>	36
3. Algorithm	36
4. Summary	37
C. Quantum support vector machines	37
1. Preliminaries	37
2. Algorithm	37
3. Summary	39
D. Quantum principal component analysis	39
1. Preliminaries	39
2. Algorithm	39
3. Summary	42
E. Quantum Neural Networks	43
1. Preliminaries	43
2. Algorithm	43
3. Summary	45
VI. Conclusions	45
References	46

I. INTRODUCTION

Machine learning is part of computer science area which aims to recognize patterns and learn from data in order to output correct predictions. It could be considered as a form of artificial intelligence supporting government analysis, medical reports, business decisions, financial risk management and other areas where decisions and optimization are based on information stored digitally. Due to an increasing amount of data stored by the companies across the world [1] and several breakthroughs in working software, machine learning is increasingly important in industry. In the last couple of years, researchers have been investigating whether quantum computers can be used to improve the performance of machine learning algorithms. A quantum computer takes advantage of quantum mechanical effects such as superposition and entanglement to solve a certain set of problems faster than classical computers. Even though quantum computers are still at the experimental stage (with some major breakthrough made by IBM that announced to build 16 qubits processor [2] with computation resources available in the cloud), the quantum algorithms have been developed for the last two decades. The quantum algorithms involve problems such as factorization of large numbers and optimization with the latter effectively used in a speed-up of machine learning algorithms. Quantum machine learning is a relatively new area of study with the recent work on quantum versions of supervised and unsupervised algorithms. The major difficulty for a non-physicist person such as a data scientist is the requirement of quantum physics theory and scientific notation knowledge. It can create some cognitive barriers to understand benefits and limitations of the quantum algorithms. Most of articles are written for a quantum physicist, leaving little or no explanation of remarkable techniques and discoveries of the quantum algorithms and the quantum machine learning algorithms. This text aims to present and explain the quantum machine learning algorithms to a data scientist in an accessible and consistent way as well as provides an introduction for a physicist interested in the topic. Moreover, it is a review of circuit-based quantum machine learning algorithms with their benefits and limitations listed. In order to properly explain quantum machine learning to a non-physicists, the paper presents an absolute minimum of quantum theory required to understand quantum computation. Then, the quantum algorithms and the quantum machine learning algorithms are presented in a step by step manner with accessible examples. Hopefully, this will allow an ambitious data scientist to understand the possibilities and limitations associated with the use of quantum computers in machine learning as well as gain knowledge about the mathematics and logic behind quantum machine learning algorithms.

II. BASIC QUANTUM THEORY

This section summarizes basic concepts of quantum theory required to understand how quantum algorithms work and how quantum effects can be used in the machine learning algorithms. Quantum mechanics seems to be counter-intuitive due to the fact that the classical world observed by our senses is too macroscopic in order to notice the rules of the microscopic world. The brief introduction to quantum theory aims to explain the quantum realm to a data scientist in a sufficient way to understand how it is used to speed-up machine learning algorithms. As a good starter, the beginnings of quantum mechanics are explained.

Before the twentieth century, physicists used to hold the view that the matter surrounding us could be either made up of tiny particles or consists of waves. The light was viewed as an electromagnetic wave that *just like* a surface water wave or a sound wave can interfere with itself making the wave amplitude greater or lower. On the other hand, the matter was attributed to be made up of particles. This view was falsified by series [3–5] of experiments, among others the photoelectric effect. The photoelectric effect assumes that electrons can be ejected from the surface of metal plate when light shines on it. According to classical electromagnetic theory increasing light amplitude would increase the kinetic energy of emitted photoelectrons, while increasing the frequency would increase measured current. Interestingly, this was contrary to experimental observations. The effect was explained by Albert Einstein [5] as he assumed that the light is a collection of particles called photons and it earned him the Nobel Prize in 1921. As it turns out later all matter in the universe manifests both particle-like and wave-like behavior, which provided the basis for a new exciting area of human knowledge, namely quantum mechanics. This is a somewhat counter-intuitive approach to the reality, so in the next sections, we will try to explain it with a big help of mathematics.

A. Quantum states

Imagine a set of positions $\{x_1, x_2, \dots, x_n\}$ in which a particle can be detected:

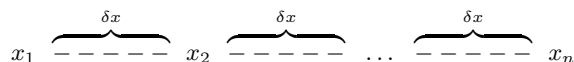


FIG. 1: Discrete set of positions in which a particle can be detected.

The gap between the positions δx is very small to provide good approximation of continuous line. The current set-up can be represented by states corresponding to a particle being detected at given locations. The state corresponding to a particle being detected at location x_1 could be written as a column vector:

$$[1, 0, \dots, 0]^T. \quad (1)$$

The next state corresponding to a particle being detected at location x_2 is associated with a column vector which is orthogonal to the previous one:

$$[0, 1, \dots, 0]^T. \quad (2)$$

Following the logic we would expect that each of these states would be an unit vector e_i in the standard basis. In terms of classical mechanics the current description is complete. However, the quantum mechanics is not so boring. In quantum realm this set-up is described by a quantum state. The quantum states are denoted using Dirac notation $|a\rangle$ which is equivalent to a n dimensional vector with complex numbers as entries. Although, we can use a standard vector notation, it is just easier to follow what other physicists do. The state corresponding to a particle being detected at position x_i is denoted as $|x_i\rangle$. The essence of quantum mechanics is that a particle before detection could be at state which is a mixture of states $\{|x_1\rangle, |x_2\rangle, \dots, |x_n\rangle\}$. Mathematically, the mixture is denoted by an arbitrary state $|\psi\rangle$ which is a linear combination of the basis states $\{|x_1\rangle, |x_2\rangle, \dots, |x_n\rangle\}$ weighted by complex amplitudes $\{c_1, c_2, \dots, c_n\}$:

$$\begin{aligned} |\psi\rangle &= c_1 |x_1\rangle + c_2 |x_2\rangle + \dots + c_n |x_n\rangle \\ &= c_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \dots + c_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}. \end{aligned} \quad (3)$$

It is simple as it is: quantum states are denoted using ket notation $|a\rangle$, which is a column vector in \mathbb{C}^n complex space (quantum states are also defined on a space with infinite number of dimensions, but this is not relevant to the quantum algorithms presented in the text). However, several questions arise:

- What does it exactly mean that the state of a particle is a linear combination?
- Why we use complex numbers as weights?

We will try to address and answer the following questions, revealing the beauty of quantum mechanics. The quantum state being in a linear combination is equivalent to say that a particle is in a quantum superposition of the basis states. The state $|\psi\rangle$ means that a particle before measurement is a *probability wave* corresponding to the probabilities of being measured at $\{x_1, x_2, \dots, x_n\}$ locations. Thus, a question *where is a particle before measurement?* according to Copenhagen interpretation of quantum mechanics is meaningless. In macroscopic terms matter seems to be localized and we could not see an item being in many places at the same time. This is due to the fact that unbelievably large number of particles is interacting with themselves so that the information from quantum states is transferred to the environment, making the quantum behavior *lost* in the jungle of interacting particles. This process is called decoherence and results in leaking of quantum nature of system to the environment. Thus, we cannot see something being *before measurement*, as even our brains are part of this interacting system. In quantum scales when we analyze behavior of a single or a few particles in isolation, the quantum effects arise. The nature of quantum mechanics is confirmed by large amount of experiments, so that this fact needs to be believed and incorporated. The considerations on the border of physics and philosophy about what happens with particle before measurement, although very interesting, are not required to understand quantum machine learning algorithms (see Bohr-Einstein debates in Ref. [6]). For us it is sufficient to say that a particle before measurement consists of different probabilities corresponding to different outcomes obtained after measurement.

Returning to the quantum state $|\psi\rangle$ note that there are many possible superpositions controlled by the value of weights $\{c_1, c_2, \dots, c_n\}$. These weights are called probability amplitudes and are strictly connected to what happens with a particle after measurement. The norm square of complex number $|c_i|^2$ gives us the probability of finding a particle in the state $|x_i\rangle$ after measurement. Due to the fact that we deal with probabilities, the probability amplitudes should be properly normalized so that:

$$|c_1|^2 + |c_2|^2 + \dots + |c_n|^2 = \sum_{i=1}^n |c_i|^2 = 1. \quad (4)$$

As an example we have an arbitrary state:

$$|\psi\rangle = c_1 |x_1\rangle + c_2 |x_2\rangle = \frac{i}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{\sqrt{3}}{2} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{i}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}. \quad (5)$$

What is the probability that a particle after measurement will be in the state $|x_1\rangle$? The answer implies taking norm square of c_1 :

$$|c_1|^2 = c_1^* c_1 = \frac{-i}{2} \frac{i}{2} = \frac{-i^2}{2^2} = \frac{1}{4}. \quad (6)$$

This can be also calculated in a more systematic way. We claim that the probability of finding a particle $|\psi\rangle$ in a state $|\phi\rangle$ after measurement is expressed as:

$$|\langle\phi|\psi\rangle|^2 \quad (7)$$

where bra notation $\langle a|$ represents a conjugate transpose of ket $|a\rangle$. The bracket $\langle a|a\rangle$ is an inner product of two vectors, which is just a number. To see how it works we calculate the inner product from the example:

$$\begin{aligned} \langle x_1|\psi\rangle &= \langle x_1| c_1 |x_1\rangle + \langle x_1| c_2 |x_2\rangle = c_1 \langle x_1|x_1\rangle + c_2 \langle x_1|x_2\rangle \\ &= \frac{i}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{\sqrt{3}}{2} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{i}{2} \end{aligned} \quad (8)$$

and then the norm square is:

$$|\langle\phi|\psi\rangle|^2 = \langle\phi|\psi\rangle^* \langle\phi|\psi\rangle = \frac{-i}{2} \frac{i}{2} = \frac{1}{4} \quad (9)$$

yielding the expected result. The orthogonality of the basis states could be interpreted in terms of probability. The probability of measuring the particle $|x_1\rangle$ to be in the state $|x_1\rangle$ always yields $|\langle x_1|x_1\rangle|^2 = 1$, whereas we would never find the same particle being in the state $|x_2\rangle$ as $|\langle x_2|x_1\rangle|^2 = 0$.

We have stated that particles manifests wave-like behavior and promised to explain that fact mathematically. The wave-like interference is fully explained by the presence of complex numbers in probability amplitudes. The probabilities in real numbers when added are always greater or equal: $p_1 + p_2 \geq p_1$ and $p_1 + p_2 \geq p_2$. The complex amplitudes when squared are also real, but now the addition of complex numbers $|c_1 + c_2|^2$ can increase or decrease the probability. The probability amplitude $c_1 = \frac{i}{2}$ when squared is equal to probability $|c_1|^2 = \frac{1}{4}$. The probability amplitude $c_2 = \frac{-i}{2}$ when squared is also equal to probability $|c_2|^2 = \frac{1}{4}$, however the sum of probability amplitudes $c_1 + c_2$ yields probability $|c_1 + c_2|^2 = |\frac{-i+i}{2}|^2 = 0$ which is certainly lower. The complex numbers can cancel or overlay each other, which has a physical meaning of interference. This is the core of quantum mechanics allowing to explain wave-like behavior of particles.

B. Quantum observables

The physics is all about measuring and analyzing quantities such as position, momentum or energy. These quantities are called observables and can be retrieved from the current state of a system. In classical physics observable F is a function that takes state S and outputs real number x which corresponds to measured quantity, that is:

$$F(S) = x. \quad (10)$$

As an example consider we measure the observable which is the heat of the gas and want to output the energy E of the system. The state S is characterized by the temperature of the system T and there exists a real-valued function F that allows us to calculate energy, so that $F(T) = E$. In quantum physics an observable is not a real-valued function, but is represented by a matrix O that acts on a quantum state $|\psi\rangle$. Just like the function F in classical physics, the matrix O allows us to retrieve quantity from a system, however in quantum physics the result of measurement is discretized. The eigenvalues λ_i of the matrix O are the only possible values observable can take after being measured. The eigenvectors $|a_i\rangle$ can be interpreted as states in which the system is left after measuring the associated eigenvalue λ_i . This is written as:

$$O|\psi\rangle \rightarrow \lambda_i|a_i\rangle \quad (11)$$

with the arrow representing measurement and collapse of the state superposition $|\psi\rangle$ to the state $|a_i\rangle$. Each quantity we want to retrieve from quantum state is associated with a different observable. In case we would like to measure position and momentum having state $|\psi\rangle$, the observables corresponding to the position and momentum are represented by different matrices. Each of these matrices has eigenvalues that are the only possible values of measured quantity, that is either position or momentum. As an example of observable, we take the position from the set-up presented in Fig. (1). A value measured is the position of a particle on the line. We know that each measurement can yield position x_i with different probabilities $|c_i|^2$. The state after the measurement is one of the $|x_i\rangle$ basis states. As a result we are looking for a matrix which eigenvalues are the real numbers x_i and associated eigenvectors are exactly $|x_i\rangle$:

$$O|\psi\rangle = \sum_{i=1}^n x_i c_i |x_i\rangle \rightarrow x_i |x_i\rangle. \quad (12)$$

Due to the fact that the eigenvalues of the matrix must be real (we do not observe in the universe position or momentum equal $1 + i$), O must be a Hermitian matrix. The searched matrix has a form of:

$$O = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_n \end{bmatrix}. \quad (13)$$

The matrix is Hermitian, the eigenvalues are obviously x_i and the corresponding eigenvectors are $|x_i\rangle$.

The distribution of possible outcomes λ_i is governed by probabilities $|c_i|^2$. Making multiple measurements of a particle in the same state $|\psi\rangle$ we could be interested in an expectation value of the observable O . Note that the phrase *multiple measurements* means that the state is prepared, measured and then prepared from scratch again. In

case we measure the same state immediately after previous measurement we would simply get collapsed state with 100% probability. An expectation value of the observable O is denoted as $\langle O \rangle$ and can be calculated as follows:

$$\begin{aligned}
\langle O \rangle &= \langle \psi | O | \psi \rangle \\
&= (c_1^* \langle a_1 | + c_2^* \langle a_2 | + \cdots + c_n^* \langle a_n |) (c_1 O | a_1 \rangle + c_2 O | a_2 \rangle + \cdots + c_n O | a_n \rangle) \\
&= (c_1^* \langle a_1 | + c_2^* \langle a_2 | + \cdots + c_n^* \langle a_n |) (c_1 \lambda_1 | a_1 \rangle + c_2 \lambda_2 | a_2 \rangle + \cdots + c_n \lambda_n | a_n \rangle) \\
&= \lambda_1 |c_1|^2 + \lambda_2 |c_2|^2 + \cdots + \lambda_n |c_n|^2
\end{aligned} \tag{14}$$

with the last line in Eq. (14) being a statistical definition of expectation value:

$$E(\lambda) = \sum_{i=1}^n \lambda_i p_i. \tag{15}$$

In these calculations we have used the fact that the basis states $|a_i\rangle$ and $|a_j\rangle$ are orthogonal for $i \neq j$ meaning that $\langle a_j | a_i \rangle = 0$ and the basis states are actually the eigenvectors of matrix O meaning that $O |a_i\rangle = \lambda_i |a_i\rangle$. As an example we calculate an expectation value of position for state:

$$|\psi\rangle = \frac{i}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{\sqrt{3}}{2} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{i}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}. \tag{16}$$

The possible positions measured are $x_1 = 1$ and $x_2 = 2$ so that the observable is represented by matrix:

$$O = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}. \tag{17}$$

The expectation value $\langle O \rangle$ is calculated in a following way:

$$\begin{aligned}
\langle O \rangle &= \langle \psi | O | \psi \rangle \\
&= \begin{bmatrix} \frac{-i}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{i}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \\
&= 1 * \frac{1}{4} + 2 * \frac{3}{4} = \frac{7}{4} = 1.75.
\end{aligned} \tag{18}$$

We might be also interested in the variance of outcomes, i.e. the spread of the possible results around the expectation value. The variance $Var(O)$ is defined as:

$$\begin{aligned}
Var(O) &= \langle \psi | (O - \langle O \rangle)^2 | \psi \rangle \\
&= (c_1^* \langle a_1 | + \cdots + c_n^* \langle a_n |) (c_1 (O - \langle O \rangle)^2 | a_1 \rangle + \cdots + c_n (O - \langle O \rangle)^2 | a_n \rangle) \\
&= (c_1^* \langle a_1 | + \cdots + c_n^* \langle a_n |) (c_1 (\lambda_1 - \langle O \rangle)^2 | a_1 \rangle + \cdots + c_n (\lambda_n - \langle O \rangle)^2 | a_n \rangle) \\
&= (\lambda_1 - \langle O \rangle)^2 |c_1|^2 + \cdots + (\lambda_n - \langle O \rangle)^2 |c_n|^2
\end{aligned} \tag{19}$$

which is a statistical definition of variance:

$$Var(\lambda) = \sum_{i=1}^n (\lambda_i - \bar{\lambda})^2 p_i \tag{20}$$

with $\bar{\lambda}$ denoting mean of the data. The variance of the observable O acting on the state $|\psi\rangle$ from the previous example in which $\langle O \rangle = 1.75$ is calculated as follows:

$$\begin{aligned}
Var(O) &= \langle \psi | (O - \langle O \rangle)^2 | \psi \rangle \\
&= \begin{bmatrix} \frac{-i}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} (1 - 1.75)^2 & 0 \\ 0 & (2 - 1.75)^2 \end{bmatrix} \begin{bmatrix} \frac{i}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \\
&= 0.5625 * \frac{1}{4} + 0.0625 * \frac{3}{4} = 0.1875
\end{aligned} \tag{21}$$

resulting in standard deviation of approximately $\sqrt{0.1875} \approx 0.433$.

C. Measurement

The previous section provided some intuition and basic facts about measurement in quantum mechanics. This section is to summarize the topic and provide the comparison to classical measurement. In classical physics the measurement is characterized by two assumptions:

- Theoretically, the measurement leaves the system in the same state as it was before,
- The result of measurement is predictable. It means that if the experiment is to be repeated, we would anticipate exactly the same outcome.

In quantum scales, these assumptions turned out to be wrong. The measurement in quantum mechanics is:

- Irreversible operation that transforms the general state $|\psi\rangle$ into an eigenvector $|a_i\rangle$ of measured observable O . It is also said that the state $|\psi\rangle$ has collapsed to the eigenvector $|a_i\rangle$,
- The result of measurement is uncertain and is always one of the eigenvalue λ_i of measured observable O . The eigenvalue will be measured with probability $|\langle a_i|\psi\rangle|^2$, where $|a_i\rangle$ is the eigenvector corresponding to eigenvalue λ_i .

Two facts emerge from the points listed. One is that if we perform measurement immediately after the first measurement we would observe state $|a_i\rangle$ with 100% probability. It is caused by the fact that the system has already collapsed to one of its eigenvectors and measurement of eigenvector always yields the same eigenvector. The second fact is related to the order of different observables measurement. The measurement is an irreversible operation, thus performing measurement of first observable will have an impact on the measurement of second observable immediately after the first measurement.

In quantum physics, there exists intriguing fact about the measurement of two observables. This fact is a well-known Heisenberg uncertainty principle which states that product of the variances of two observables is always greater than a threshold equal to one-fourth of squared expected value of their commutator:

$$Var(O_1)Var(O_2) \geq \frac{1}{4} |\langle [O_1, O_2] \rangle|^2 . \quad (22)$$

The commutator is defined as a difference:

$$[O_1, O_2] = O_1O_2 - O_2O_1 \quad (23)$$

which for matrices is not necessarily zero. In order to fully understand the Heisenberg uncertainty principle we analyze the following example:

- Prepare k repetitions of experiment (exactly the same states are measured in each experiment),
- Measure position and then momentum in each experiment,
- Write the results and calculate the variance for both position and momentum outcomes.

It turns out that the product of these variances will be always greater than a non-zero threshold because the Hermitian matrices corresponding to position and momentum do not commute $[O_1, O_2] \neq 0$. The exact value of the threshold is $\frac{\hbar^2}{4}$ where \hbar is a reduced Planck's constant, however the numeric value is not required to understand the principle. The non-commutativity is interpreted as the fact that observables O_1 and O_2 do not share the eigenvectors. In case O_1 and O_2 commute they have identical eigenvectors $|a_i\rangle$ corresponding to some eigenvalues λ_i and μ_i so that:

$$\begin{cases} O_1 |a_i\rangle &= \lambda_i |a_i\rangle \\ O_2 |a_i\rangle &= \mu_i |a_i\rangle . \end{cases} \quad (24)$$

Thus the second observable O_2 can be measured without disturbing already collapsed eigenvector $|a_i\rangle$. Some of the sources explain that the product of variances is greater than some threshold for non-commuting observables because the measurement itself disturbs the state. Although the effects of disturbing are not negligible, this is not the true reason why the Heisenberg uncertainty principle holds. The principle arises from the fact that particles are waves, not point objects and it is a fundamental property of the universe, so that even if the measurement is done without any disturbance of the system (obviously except collapsing of state) the uncertainty for non-commuting observables still exists.

D. Assembling quantum states

Assume we want to analyze multi-particle states instead of only one particle states. The machinery that should be used to accomplish that is called tensor product of state spaces and the procedure is called assembling of quantum states. Having k independent particle states: $\{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_k\rangle\}$ we can describe them by general state Ψ :

$$|\Psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \dots \otimes |\psi_k\rangle = |\psi_1\psi_2\dots\psi_k\rangle. \quad (25)$$

Thus, if $|\psi_i\rangle$ are n -dimensional vectors for each i , then the state $|\Psi\rangle$ will have n^k dimensions. As an example, let us add another particle to the example presented in Fig. (1), so that the system is illustrated by:

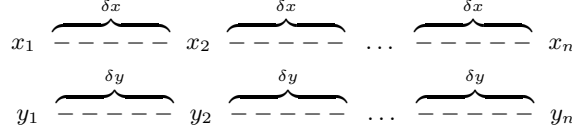


FIG. 2: Two discrete sets of positions in which particle x and y can be detected.

Now, the possible states are the combinations of x particle position and y particle position. The general state $|\Psi\rangle$ for $n = 2$ possible outcomes is constructed as follows:

$$\begin{aligned} |\Psi\rangle &= |\psi_x\rangle \otimes |\psi_y\rangle \\ &= (c_{1,x}|x_1\rangle + c_{2,x}|x_2\rangle) \otimes (c_{1,y}|y_1\rangle + c_{2,y}|y_2\rangle) \\ &= \overbrace{c_{1,x}c_{1,y}}^{c_1}|x_1y_1\rangle + \overbrace{c_{1,x}c_{2,y}}^{c_2}|x_1y_2\rangle + \overbrace{c_{2,x}c_{1,y}}^{c_3}|x_2y_1\rangle + \overbrace{c_{2,x}c_{2,y}}^{c_4}|x_2y_2\rangle \\ &= c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= c_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \end{aligned} \quad (26)$$

where we implicitly assume that the outcomes of particle measurement are described by eigenvectors $e_1 = [1, 0]^T$ and $e_2 = [0, 1]^T$. The state $|\Psi\rangle$ is $2^2 = 4$ dimensional vector and is in a superposition of all possible position measurement outcomes of two particles. The interpretation of complex numbers c_i is exactly the same as in one particle states, for instance c_2 is a probability amplitude for measuring the first particle at position x_1 and the second particle at position y_2 .

The assembling of quantum states is strictly related to another astonishing property of quantum theory, that is quantum entanglement. This concept will be explained based on a simple example. Assume existence of two quantum states:

$$|\psi\rangle = \frac{1}{2}|aa\rangle + \frac{1}{2}|ba\rangle + \frac{1}{2}|ab\rangle + \frac{1}{2}|bb\rangle \quad (27)$$

and

$$|\phi\rangle = \frac{1}{\sqrt{2}}|aa\rangle + \frac{1}{\sqrt{2}}|bb\rangle. \quad (28)$$

Both states contains two particles in a superposition of states $|a\rangle$ and $|b\rangle$. The following table presents the possible outcomes with corresponding probabilities for state $|\psi\rangle$:

TABLE I: Possible outcomes for state $|\psi\rangle$

	Particle 2: $ a\rangle$	Particle 2: $ b\rangle$
Particle 1: $ a\rangle$	$ \psi\rangle \rightarrow aa\rangle: \frac{1}{4}$	$ \psi\rangle \rightarrow ab\rangle: \frac{1}{4}$
Particle 1: $ b\rangle$	$ \psi\rangle \rightarrow ba\rangle: \frac{1}{4}$	$ \psi\rangle \rightarrow bb\rangle: \frac{1}{4}$

As an example measuring first particle yields state $|a\rangle$ with probability $\frac{1}{4} + \frac{1}{4} = 0.5$. The next table illustrates the outcomes of measuring two particles in the second state $|\phi\rangle$:

TABLE II: Possible outcomes for state $|\phi\rangle$

	Particle 2: $ a\rangle$	Particle 2: $ b\rangle$
Particle 1: $ a\rangle$	$ \phi\rangle \rightarrow aa\rangle: \frac{1}{2}$	$ \phi\rangle \rightarrow ab\rangle: 0$
Particle 1: $ b\rangle$	$ \phi\rangle \rightarrow ba\rangle: 0$	$ \phi\rangle \rightarrow bb\rangle: \frac{1}{2}$

Measuring the first particle yields state $|a\rangle$ with probability 0.5, however now the measurement immediately collapsed the second particle state to the state $|a\rangle$. The individual states of two particles are related to each other and this relation is called quantum entanglement. The most astonishing part of quantum entanglement is that it does not depend on the distance. Assume we prepare two particles to be entangled and send the second one to other galaxy (without disturbing it). Then observer A measures the first particle on Earth and gets state $|a\rangle$. The observer B must also get state $|a\rangle$ measuring the second particle in other galaxy. The measurement outcome of the first particle will always immediately determine measurement outcome of the second particle. It does not mean that the information is moving faster than light, because to communicate results between observer A and observer B the classical communication channel still needs to be used, however once the results are compared we will always find out that the effects of quantum entanglement hold. The state is in quantum entanglement if it cannot be rewritten as a tensor product of single particle states. As an example, the state $|\psi\rangle$ could be rewritten as:

$$\begin{aligned} |\psi\rangle &= \left(\frac{1}{\sqrt{2}} |a\rangle + \frac{1}{\sqrt{2}} |b\rangle \right) \otimes \left(\frac{1}{\sqrt{2}} |a\rangle + \frac{1}{\sqrt{2}} |b\rangle \right) \\ &= \frac{1}{2} |aa\rangle + \frac{1}{2} |ba\rangle + \frac{1}{2} |ab\rangle + \frac{1}{2} |bb\rangle . \end{aligned} \quad (29)$$

For entangled state $|\phi\rangle$ this is not possible, meaning that the states are not separable. To prove this assume there exist complex numbers c_1, c_2 and c'_1, c'_2 so that:

$$|\phi\rangle = (c_1 |a\rangle + c_2 |b\rangle) \otimes (c'_1 |a\rangle + c'_2 |b\rangle) = \frac{1}{\sqrt{2}} |aa\rangle + \frac{1}{\sqrt{2}} |bb\rangle . \quad (30)$$

However, this would imply that $c_1 c'_1 = c_2 c'_2 = \frac{1}{\sqrt{2}}$ and $c_1 c'_2 = c_2 c'_1 = 0$, which obviously does not have any solution. Thus, the state $|\phi\rangle$ cannot be rewritten as a tensor product of single particle states.

E. Density matrix

There are some cases that states machinery is not a sufficient tool to describe a quantum system. In quantum mechanics we are considering two forms of probabilities :

- One that relates to the states and means probability of measuring an arbitrary state in one of eigenvectors of an observable O . The probability is derived from the norm square of probability amplitude c_i . This form is associated with pure states,
- The other form relates to the whole system and means the probability of finding the particular system in one of the possible states. The probability expresses the fact that we do not have a full knowledge about the system and we can only say that the system is in i -th pure state with probability p_i . This form is associated with mixed states.

In order to describe both pure and mixed states a density matrix notation is used. A density matrix for pure state is defined as:

$$\rho = |\psi\rangle \langle\psi| \quad (31)$$

and for mixed state as:

$$\rho = \sum_{i=1}^n p_i |\psi_i\rangle \langle\psi_i| \quad (32)$$

where p_i is interpreted as finding mixed state in the state $|\psi_i\rangle$. Actually, the density matrix for pure state is just a special case of density matrix for mixed state with $p_1 = 1$. The expression $|\psi\rangle\langle\psi|$ is calculated as tensor product. To see how it is done we analyze the following pure state:

$$|\psi\rangle = \frac{i}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{\sqrt{3}}{2} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{i}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \quad (33)$$

then we take the tensor product to express density matrix:

$$\begin{aligned} \rho = |\psi\rangle\langle\psi| &= \begin{bmatrix} \frac{i}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \otimes \begin{bmatrix} \frac{-i}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{i}{2} \frac{-i}{2} & \frac{i}{2} \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} \frac{-i}{2} & \frac{\sqrt{3}}{2} \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{i\sqrt{3}}{4} \\ \frac{-i\sqrt{3}}{4} & \frac{3}{4} \end{bmatrix}. \end{aligned} \quad (34)$$

From Eq. (18) in the previous section we know that the expectation value of state $|\psi\rangle$ given an observable O :

$$O = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad (35)$$

is equal to 1.75. It turns out that the same result could be obtained using density matrix and following formula:

$$\begin{aligned} \langle O \rangle &= \text{Tr}(\rho O) \\ &= \text{Tr} \left(\begin{bmatrix} \frac{1}{4} & \frac{i\sqrt{3}}{4} \\ \frac{-i\sqrt{3}}{4} & \frac{3}{4} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \right) = \text{Tr} \left(\begin{bmatrix} \frac{1}{4} & \frac{2i\sqrt{3}}{4} \\ \frac{-i\sqrt{3}}{4} & \frac{6}{4} \end{bmatrix} \right) \\ &= \frac{1}{4} + \frac{6}{4} = 1.75 \end{aligned} \quad (36)$$

yielding result as expected. As an example of mixed state, assume the situation in which we are unsure whether the system has been prepared in state $|\psi\rangle$ or $|\phi\rangle$, however we know the probability of system being in the first state is $p = \frac{1}{4}$ and being in the second state is $q = 1 - p = \frac{3}{4}$. The additional state $|\phi\rangle$ is defined as:

$$|\phi\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}. \quad (37)$$

The expected value of this state given the same observable O is 1.5, as in half of the cases we will find the state in the first eigenvector with eigenvalue $\lambda_1 = 1$ and in half of the cases we will find the state in the second eigenvector with eigenvalue $\lambda_2 = 2$. It is not possible to represent a mixed state using state formalism, however it could be described with the density matrix:

$$\begin{aligned} \rho &= p |\psi\rangle\langle\psi| + q |\phi\rangle\langle\phi| = \\ &= \frac{1}{4} \begin{bmatrix} \frac{1}{4} & \frac{i\sqrt{3}}{4} \\ \frac{-i\sqrt{3}}{4} & \frac{3}{4} \end{bmatrix} + \frac{3}{4} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{7}{16} & \frac{6+i\sqrt{3}}{16} \\ \frac{6-i\sqrt{3}}{16} & \frac{9}{16} \end{bmatrix}. \end{aligned} \quad (38)$$

For the mixed state we anticipate that in 25% of outcomes the average will be 1.75 and in 75% of outcomes the average will be 1.5, thus yielding expected value of 1.5625. This result could be also obtained using exactly the same formula Eq. (36) as for pure states:

$$\begin{aligned} \langle O \rangle &= \text{Tr}(\rho O) \\ &= \text{Tr} \left(\begin{bmatrix} \frac{7}{16} & \frac{6+i\sqrt{3}}{16} \\ \frac{6-i\sqrt{3}}{16} & \frac{9}{16} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \right) = \frac{7}{16} + \frac{18}{16} = 1.5625. \end{aligned} \quad (39)$$

Thus, the density matrix is a common tool for describing both pure and mixed states and allows us to use quantum mechanics in case we do not have full knowledge about the system. For further information about density matrices and other concepts in basic quantum mechanics see Ref. [7].

III. QUANTUM COMPUTATION

A. Qubit

The definition of qubit lies in the center of quantum computation theory. Before we dive into the explanation of qubit, let us recall the definition of classical bit. A bit is an unit of information, which describes a two-dimensional classical system. Thus, the classical system could be either in the state:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (40)$$

or in the state:

$$|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (41)$$

The physical representation of a bit is two *flip-flop* states representation, for instance two distinct voltages of electric circuit or two distinct levels of light intensity. This is sufficient for classical physics and this how the classical computer works. The quantum computer uses the effects of quantum mechanics such as a superposition of states. A qubit is an unit of information, which describes a two-dimensional quantum system and the general state of qubit is represented by a pair of complex numbers:

$$c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad (42)$$

so that it is a superposition of the states $|0\rangle$ and $|1\rangle$. The physical representation of qubit could be the polarization of a photon, spin of a particle or ground and excited orbit of an electron in atom. A classical computer handles with a string of bits for instance 01010001 and based on it does the calculations. Quantum computation assumes that qubits can be assembled using tensor product, thus the same string could be written as:

$$\begin{aligned} & |0\rangle \otimes |1\rangle \otimes |0\rangle \otimes |1\rangle \otimes |0\rangle \otimes |0\rangle \otimes |0\rangle \otimes |1\rangle \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned} \quad (43)$$

which is a vector with 256 rows:

$$|01010001\rangle = \begin{matrix} |00000000\rangle \\ |00000001\rangle \\ \vdots \\ |01010000\rangle \\ |01010001\rangle \\ \vdots \\ |11111110\rangle \\ |11111111\rangle \end{matrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}. \quad (44)$$

Note that the general state of 8 qubit quantum computer could be written as a superposition:

$$\begin{aligned} |\psi\rangle &= c_1 |00000000\rangle + \cdots + c_{82} |01010001\rangle + \cdots + c_{255} |11111110\rangle + c_{256} |11111111\rangle \\ &= \sum_{i=1}^N c_i |i\rangle \end{aligned} \quad (45)$$

where a set of $|i\rangle$ is called the standard computational basis and complex numbers are normalized, so that $\sum_{i=1}^N |c_i|^2 = 1$. This illustrates the overwhelming difference between quantum and classical computers. In order to write 8 qubit system 256 complex numbers are required, whereas on classical computer only 8 zeros or ones are required to fully describe 8 bit system. The effect increases exponentially with number of qubits, for 64 qubits we will need $2^{64} = 18,446,744,073,709,551,616$ complex numbers to emulate quantum state on a classical machine. The quantum algorithms are exploiting this astonishing fact as we will see in the further sections.

B. Quantum gates

A classical logical gate is a way of bits manipulation. As an example the gate *NOT* flips the bit so that $NOT|0\rangle = |1\rangle$ and $NOT|1\rangle = |0\rangle$. This can be represented by following 2-by-2 matrix:

$$NOT = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (46)$$

The matrix defined in Eq. (46) satisfies:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (47)$$

and

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (48)$$

The second example of a classical gate is *AND* gate, which accepts two bits and outputs one. The *AND* gate is represented by 2-by-4 matrix:

$$AND = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (49)$$

and satisfies the following relations:

- $AND|00\rangle = |0\rangle$,
- $AND|01\rangle = |0\rangle$,
- $AND|10\rangle = |0\rangle$,
- $AND|11\rangle = |1\rangle$.

Quantum gates are a way of qubits manipulation. A state enters a gate in quantum circuit and exits as other state, thus quantum gates represent time evolution of a state describing qubits. The quantum gate satisfies the following criteria:

- must preserve norms i.e. norm squared probability amplitudes sum to one after gate application,
- must be reversible i.e. evolution of each not measured quantum state must be reversible.

These conditions are equivalent to a restriction that quantum gates must be unitary matrices. Thus, the 2x4 AND gate is not a valid quantum gate, because it is not unitary, whereas NOT gate is a valid quantum gate also known as one of the three Pauli matrices used in quantum mechanics:

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (50)$$

Another trivial example of unitary matrix is the identity matrix:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (51)$$

Frequently used gate in quantum computation is Hadamard gate which allows to produce superposition of states:

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (52)$$

Applying Hadamard gate on qubit in state $|0\rangle$ yields:

$$H|0\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \quad (53)$$

and on qubit in state $|1\rangle$:

$$H|1\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{|0\rangle - |1\rangle}{\sqrt{2}}. \quad (54)$$

There are also quantum gates acting on two qubits, for instance *SWAP* gate represented by:

$$SWAP = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (55)$$

It swaps two qubits so that the example state $|01\rangle$ is evolved into state $|10\rangle$:

$$SWAP|01\rangle = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = |10\rangle. \quad (56)$$

We can apply quantum gates on more than one qubit using tensor product. As an example we initialize two qubits in state $|\psi_0\rangle = |00\rangle$, apply Hadamard gate only on the first qubit leaving the second qubit unchanged and resulting in state $|\psi_1\rangle$:

$$\begin{aligned} |\psi_1\rangle &= (H \otimes I) |00\rangle \\ &= \frac{1}{\sqrt{2}} \left(\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} (|00\rangle + |10\rangle) = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |0\rangle. \end{aligned} \quad (57)$$

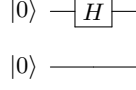


FIG. 3: The operation $|\psi_0\rangle \rightarrow |\psi_1\rangle$ presented as quantum circuit.

In general we can combine any number of quantum gates using tensor products and use it on qubit state as long as the dimensions of the state vector matches the size of combined quantum gate. The operation $|\psi_0\rangle \rightarrow |\psi_1\rangle$ can be presented in quantum circuit:

Continuing the example, the state $|\psi_1\rangle$ could be evolved to $|\psi_2\rangle$ using *SWAP* gate on two qubits in a superposition:

$$\begin{aligned}
 |\psi_2\rangle &= \text{SWAP} \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes |0\rangle \right) \\
 &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \\
 &= \frac{1}{\sqrt{2}} (|00\rangle + |01\rangle) = |0\rangle \otimes \frac{|0\rangle + |1\rangle}{\sqrt{2}}
 \end{aligned} \tag{58}$$

As expected the *SWAP* gate interchanged the qubits. This operation can be presented in quantum circuit in Fig. (4).

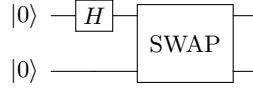


FIG. 4: The operation $|\psi_0\rangle \rightarrow |\psi_2\rangle$ presented as quantum circuit.

The last step is to measure both qubits. Measurement is an irreversible operation, thus is not a quantum gate, but a final step of quantum circuit. Measuring the first qubit we will always yield state $|0\rangle$ and measuring the second qubit we have 50% chances of getting qubit in state $|0\rangle$ and 50% chances of getting qubit in state $|1\rangle$. The measurement can be illustrated in Fig. (5) using meter symbol.

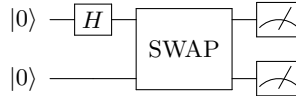


FIG. 5: The operation $|\psi_0\rangle \rightarrow |\psi_2\rangle$ and measurement presented as quantum circuit.

The quantum computation can be defined as a unitary evolution of a closed quantum system - it takes some initial state as an input and outputs a final state, which can be measured to retrieve specific information. It is crucial that a quantum system remains isolated, i.e. there is no decoherence. Any decoherence is a loss of information from a system into the environment and relevant quantum benefits can vanish. Interestingly, the quantum computation using small number of qubits can be simulated on classical computer with a few software solutions being open-source [8, 9].

C. Quantum parallelism

The quantum speed-up of algorithm comes from quantum parallelism achieved through the superposition of qubits. The quantum parallelism is a feature of many quantum algorithms and in simple words allows to evaluate function

on many inputs simultaneously. Suppose we have a function $f(\phi) : \{0, 1\}^n \rightarrow \{0, 1\}$. We initialize a quantum system in the state $|\psi_0\rangle = |\phi, 0\rangle$ and evolve it to the state $|\psi_1\rangle = O|\phi, 0\rangle$ with an appropriate sequence of quantum gates, which combined can be represented by gate O . The quantum gate O takes the input $|x, 0\rangle$ and evolves it to $|x, f(x)\rangle$. For simplicity, assume that $n = 2$ and our first register with state $|\phi\rangle$ is initialized as:

$$|\phi\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + |1\rangle}{\sqrt{2}} \quad (59)$$

which can be achieved by application of Hadamard gate on both qubits $|00\rangle$ (however this operation is assumed to be done before in a separate quantum circuit). The quantum circuit is presented in Fig. 6.

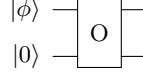


FIG. 6: Application of quantum gate O presented as quantum circuit.

To summarize we started with the state $|\psi_0\rangle$:

$$|\psi_0\rangle = |\phi, 0\rangle = \frac{1}{2} (|00\rangle + |01\rangle + |10\rangle + |11\rangle) \otimes |0\rangle \quad (60)$$

and ended with the state $|\psi_1\rangle$:

$$\begin{aligned} |\psi_1\rangle &= O|\phi, 0\rangle = \frac{1}{2} (O|00, 0\rangle + O|01, 0\rangle + O|10, 0\rangle + O|11, 0\rangle) \\ &= \frac{1}{2} (|00, f(00)\rangle + |01, f(01)\rangle + |10, f(10)\rangle + |11, f(11)\rangle) . \end{aligned} \quad (61)$$

Thus, with just one operation we evaluated function f for four inputs 00, 01, 10 and 11 simultaneously. The same operation on a classical computer must be executed one by one for each input or done in separate classical circuits. This remarkable feature of quantum algorithms uses the fact that input state is in superposition. The general logic can be expanded for large n making algorithm more effective. The only disadvantage is that the outputs of function are stored as quantum states, thus measuring the second qubit yields only one evaluated value. Nevertheless, the quantum parallelism is effectively used by many quantum algorithms as an intermediate step in quantum circuit.

IV. QUANTUM ALGORITHMS

This section presents quantum algorithms, which will be used by quantum machine learning algorithms as subroutines. The quantum algorithms presented are:

- Grover's search algorithm (basis for quantum minimization algorithm),
- Quantum minimization algorithm,
- Quantum Fourier transform algorithm (basis for quantum phase estimation algorithm),
- Quantum phase estimation algorithm.

In theory, each quantum algorithm is either quadratically or exponentially faster than its classical counterpart. To measure algorithm efficiency the time complexities are used along with big-O notation. The big-O notation explains how fast time complexity of an algorithm grows when a main element is arbitrarily large. The main element is meant to be a size of input, a number of quantum gates used or a number of iterations used depending on which factor impacts the effectiveness of algorithm most. It presents the limiting behavior of an algorithm avoiding constants and not significant terms, which are much smaller in comparison to the main element that drives the algorithm time complexity. The Fig. 7 gives the intuition about the big-O notation and indicates the order of time complexities.

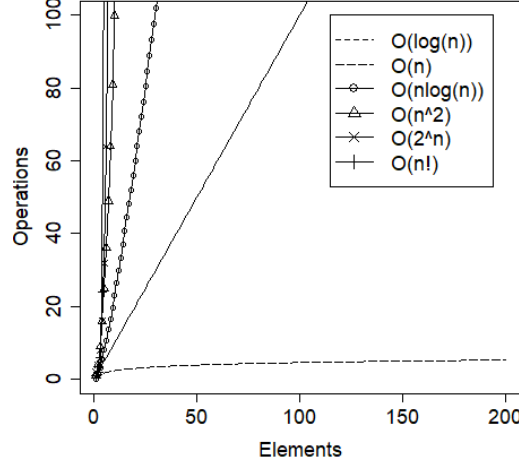


FIG. 7: Big-O notation

Each algorithm is described in the following manner: a brief introduction in preliminaries, a step by step algorithm explanation based on a simple example, then a general description of quantum algorithm and finally the conclusions about time complexity and other remarks in summary.

A. Grover's search algorithm

1. Preliminaries

Grover's algorithm is a quantum search algorithm, which runtime is quadratically faster than any classical counterpart [10]. Grover's algorithm is not intended to find an element in a database, its purpose is searching through function inputs to check whether the function returns true for that input. It is very efficient in case the function is unknown or extremely complex and we want to know for which input the function returns true or for which input the equation is solved. This function can be represented as a quantum oracle and may be built from a large number of combined quantum gates. It is assumed that the quantum oracle has been already given to the algorithm and serves as a black-box. As an example, imagine that we are given a function $f(x) : \{0, 1\}^n \rightarrow \{0, 1\}$ as a set of *AND* and *OR* logical operators which for exactly one binary string of zeros and ones returns true. Coding such a function can be relatively simple, however examining for which combination of zeros and ones the function returns true would require at worst scenario 2^n function calls on a classical machine. On a quantum computer, we can transform the function into a valid set of quantum gates constituting for a quantum oracle O and use Grover's search algorithm to find a correct input with very high accuracy in only $\sqrt{2^n}$ iterations.

2. Example

Grover's search algorithm relies on setting a state of the system into a superposition of all possible inputs and then, the probability of finding the searched input is increased in each algorithm iteration. We aim to explain the algorithm starting with a simple $n = 2$ dimensional example. As an input, the algorithm requires a quantum oracle O that *picks out* the string 10. The function representing that operation returns true if input is equal to searched binary string:

$$f(x) = \begin{cases} 1 & x = 10 \\ 0 & x \neq 10 \end{cases} . \quad (62)$$

Then we need to represent function $f(x) : \{0, 1\}^2 \rightarrow \{0, 1\}$ as quantum oracle O . This is a little bit tricky as first guess $O|x\rangle = |f(x)\rangle$ is not unitary and reversible operation, thus is not a quantum gate (the state $|x\rangle$ consists of 2 qubits, whereas $|f(x)\rangle$ of only one qubit). However, there exists a method of constructing the quantum oracle as:

$$O|x\rangle = (-1)^{f(x)} |x\rangle . \quad (63)$$

The flip of amplitude sign in case $f(x) = 1$ marks the state 10 and the matrix representing this operation is unitary, so that O is desired quantum oracle. In that particular example the quantum oracle acts on two input qubits and is expressed as matrix:

$$O = \begin{matrix} & |00\rangle & |01\rangle & |10\rangle & |11\rangle \\ \begin{matrix} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}. \quad (64)$$

At this point it is important to realize that building quantum oracle for that example has already unveiled for which input the function returns true. However, this is just for presentation purposes and in practice we have a black-box quantum oracle that is already given to the algorithm.

Obviously, the algorithm searches for the string $x = 10$, thus we require that the measurement of the system at the end should reveal the correct answer with high probability. The detailed steps of the algorithm for that particular example are explained below.

Initialize

Initialize $|0\rangle^{\otimes 2}$ qubits resulting in the state $|\psi_0\rangle = |00\rangle$.

Apply Hadamard gates

Apply Hadamard gate on each qubit resulting in the state:

$$\begin{aligned} |\psi_1\rangle &= H^{\otimes 2} |\psi_0\rangle = \frac{1}{\sqrt{2^2}}(|0\rangle + |1\rangle) \otimes (|0\rangle + |1\rangle) \\ &= \frac{1}{\sqrt{2^2}}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = \left[\frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \quad \frac{1}{2} \right]^T. \end{aligned} \quad (65)$$

After multiplication it represents all of the possible inputs to quantum oracle in equally weighted superposition. We used only 2 qubits to represent 4 inputs.

Iteration (1): 1. Apply quantum oracle O gate

Amplitude sign of the searched input is flipped with this operation resulting in the state:

$$|\psi_2^{(1)}\rangle = O |\psi_1\rangle = \frac{1}{\sqrt{2^2}}(|00\rangle + |01\rangle - |10\rangle + |11\rangle) = \left[\frac{1}{2} \quad \frac{1}{2} \quad -\frac{1}{2} \quad \frac{1}{2} \right]^T. \quad (66)$$

The quantum oracle is not enough to recognize the searched input, because the sign of amplitude has no effect on measurement probability. We have to look for additional quantum gates that makes the absolute value of amplitude greater for the searched state.

Iteration (1): 1. Inversion around the mean

In order to strengthen amplitude with negative sign we use a procedure called inversion around the mean. The operation is fairly simple and implies using $2A - I^{\otimes 2}$ gate, where A is defined as:

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad (67)$$

and the inversion around mean is represented by:

$$2A - I^{\otimes 2} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix}. \quad (68)$$

The gate has following action on the $|\psi_2\rangle$:

$$|\psi_3^{(1)}\rangle = (2A - I^{\otimes 2}) |\psi_2^{(1)}\rangle = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \quad (69)$$

The searched input has probability amplitude of one, however this is not a general case for number of input qubits greater than two. In that cases the steps: application of quantum oracle gate and application of inversion around mean are repeated around $\frac{\pi}{4}\sqrt{2^n}$ times [10]. As a quick example you can check that for 3-qubit inputs (with the second element being searched) the two iterations in the algorithm would follow:

$$\begin{aligned} |\psi_2^{(1)}\rangle &= \left[\frac{\sqrt{2}}{4} \quad -\frac{\sqrt{2}}{4} \quad \frac{\sqrt{2}}{4} \quad \frac{\sqrt{2}}{4} \quad \frac{\sqrt{2}}{4} \quad \frac{\sqrt{2}}{4} \quad \frac{\sqrt{2}}{4} \quad \frac{\sqrt{2}}{4} \right]^T \\ \rightarrow |\psi_3^{(1)}\rangle &= \left[\frac{\sqrt{2}}{8} \quad \frac{5\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \right]^T \\ \rightarrow |\psi_2^{(2)}\rangle &= \left[\frac{\sqrt{2}}{8} \quad -\frac{5\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \quad \frac{\sqrt{2}}{8} \right]^T \\ \rightarrow |\psi_3^{(2)}\rangle &= \left[\frac{\sqrt{2}}{16} \quad -\frac{11\sqrt{2}}{16} \quad \frac{\sqrt{2}}{16} \quad \frac{\sqrt{2}}{16} \quad \frac{\sqrt{2}}{16} \quad \frac{\sqrt{2}}{16} \quad \frac{\sqrt{2}}{16} \quad \frac{\sqrt{2}}{16} \right]^T \end{aligned} \quad (70)$$

so that the absolute value of amplitude for the second element is clearly higher. The equivalent representation of inversion around mean gate that is implemented in quantum circuit implies using gates:

$$2A - I^{\otimes 2} = H^{\otimes 2} (2|0^{\otimes 2}\rangle\langle 0^{\otimes 2}| - I^{\otimes 2}) H^{\otimes 2}. \quad (71)$$

Measurement

The last step implies collapse of the state:

$$|\psi_3\rangle = \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^T = 0|00\rangle + 0|01\rangle + 1|10\rangle + 0|11\rangle \quad (72)$$

by making measurement on both qubits. In that particular example the square of amplitude equals one, so that the searched input is always measured.

3. Algorithm

Generally, the algorithm could be visualized by the quantum circuit presented in Fig. (8). Grover search algorithm requires a quantum oracle O acting on n input qubits. Measurement of the system at the end of the algorithm should reveal the searched string x with high probability.

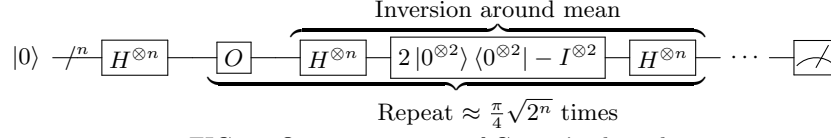


FIG. 8: Quantum circuit of Grover's algorithm.

Initialize

Initialize $|0\rangle^{\otimes n}$ qubits resulting in the state $|\psi_0\rangle = |0\dots 0\rangle$.

Apply Hadamard gates

Apply Hadamard gate on each qubit resulting in the state:

$$|\psi_1\rangle = H^{\otimes n} |\psi_0\rangle = \frac{1}{\sqrt{2^n}} (|0\dots 00\rangle + |0\dots 01\rangle + \dots + |1\dots 11\rangle). \quad (73)$$

Iteration (i): 1. Apply quantum oracle O gate

The amplitude sign of searched input is flipped with this operation resulting in the state:

$$|\psi_2^{(i)}\rangle = O |\psi_1\rangle. \quad (74)$$

Iteration (i): 2. Inversion around the mean

Apply inversion around the mean gate:

$$|\psi_3^{(i)}\rangle = (H^{\otimes n} (2|0^{\otimes n}\rangle\langle 0^{\otimes n}| - I^{\otimes n}) H^{\otimes n}) |\psi_2\rangle. \quad (75)$$

Repeat steps: Apply quantum oracle O gate and inversion around the mean about $i_{max} \approx \frac{\pi}{4}\sqrt{2^n}$ times.

Measurement

Measure the state $|\psi_3^{(i_{max})}\rangle$. The searched item should be measured with a high probability.

4. Summary

The quantum parallelism of Grover's search algorithm relies on changing amplitudes of all inputs simultaneously. This is done thanks to the superposition of states, which is purely a quantum concept. Moreover, the search is done globally, which indicates a significant improvement in optimization routines and in fact this is explored in the next section. It is known that the quantum algorithm is optimal, which means that quantum computers cannot do better in these type of problems [11]. Grover's search algorithm time complexity is around $O(\sqrt{2^n})$. In the case of a classical computer, this would be $O(2^n)$, which means a quadratic speed-up of the searching algorithm. It is not yet a holy-grail exponential speed-up, however it is still significant for large data vectors. On the other hand, Grover's algorithm is sensitive to the number of iterations. More iterations will decrease the amplitude of correct answer, thus incorrectly selecting this parameter can *overcook* the solution. Moreover, the operation of algorithm is limited in case of introducing a noise to the quantum system which is a reality in today's quantum computers [12]. There also exists extension of Grover's search algorithm for finding k matching entries instead of one [10].

B. Quantum minimization algorithm

1. Preliminaries

Grover's search algorithm is a special case of the optimization problem, as it finds an input for which $f(x) = 1$ or in other words global maximum of a function returning zeros or one. This is extended in the quantum minimization

algorithm, known also as Dürr and Høyer minimization algorithm to deal with more general problems [13]. The quantum algorithm uses Grover's search algorithm (version for k matching entries) with quantum oracles indicating which elements are smaller than an arbitrary threshold. Grover's search is performed several times in order to find a solution to the optimization problem. More importantly, thanks to a quantum parallelism global minimum is always found, which has large implications for machine learning procedures, as they often stuck in local optima.

2. Example

We would minimize an objective function $f(x) : \{0, 1\}^3 \rightarrow \{0, 1, 2, 3\}$:

$$f(x) = \begin{cases} 0 & x = 100 \\ 1 & x = 000 \\ 2 & x = 001 \\ 3 & \text{otherwise} \end{cases}, \quad (76)$$

As an input a quantum oracle O is required that *marks* inputs for which function is lower than a threshold y . For simplicity we assume that $y = \{0, 1, 2, 3\}$. The function representing that operation returns true if $f(x)$ is lower than y :

$$h(x) = \begin{cases} 1 & f(x) < y \\ 0 & f(x) \geq y \end{cases}. \quad (77)$$

Then we use a standard prescription defined in Eq. (63) to build a quantum oracle:

$$O|x\rangle = (-1)^{h(x)}|x\rangle. \quad (78)$$

This time Grover's search algorithm finds k matching entries, so that several amplitude signs may be changed. The simplest matrix representation of quantum oracle can be written as:

$$O = \delta(y-0)O_0 + \delta(y-1)O_1 + \delta(y-2)O_2 + \delta(y-3)O_3 \quad (79)$$

where $\delta(x-a)$ is Dirac delta function returning one if $x=a$ and zero otherwise. The component O_2 is written as:

$$O_2 = \begin{matrix} & \begin{matrix} |000\rangle & |001\rangle & |010\rangle & |011\rangle & |100\rangle & |101\rangle & |110\rangle & |111\rangle \end{matrix} \\ \begin{matrix} |000\rangle \\ |001\rangle \\ |010\rangle \\ |011\rangle \\ |100\rangle \\ |101\rangle \\ |110\rangle \\ |111\rangle \end{matrix} & \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \quad (80)$$

so that it flips the amplitude sign for inputs where $f(x) < 2$ i.e. $x = 100$ and $x = 000$. Other components O_0 , O_1 and O_3 are represented analogously. Similarly to Grover's search algorithm it is important to realize that building quantum oracle for that example already unveils for which input the function has minimum. However, this is just for presentation purposes and in practice we have a black-box quantum oracle that is already given to the algorithm.

The correct answer $x_{min} = 100$ for which function has global minimum should be measured with high probability. The detailed steps of the algorithm for that particular example are explained below.

Set starting point

Randomly select input x_1 and set $y_1 = f(x_1)$ as a starting point of the algorithm. In our example we randomly selected $x_1 = 001$ for which $y_1 = 2$. By expressing function outputs as binary strings:

$$\begin{aligned} 0 &\rightarrow 00 \\ 1 &\rightarrow 01 \\ 2 &\rightarrow 10 \\ 3 &\rightarrow 11 \end{aligned} \tag{81}$$

and then converting them to quantum states, a digit 2 can be represented by the state $|10\rangle$.

Iteration (1): 1. Initialize

Initialize the first register with 3 input qubits and the second register with 2 qubits to store threshold y_1 . The initial state can be written as:

$$|\psi_0^{(1)}\rangle = |000\rangle |10\rangle. \tag{82}$$

Iteration (1): 2. Apply Hadamard gates

Apply Hadamard gates on the first register resulting in the state:

$$\begin{aligned} |\psi_1^{(1)}\rangle &= (H^{\otimes 3} \otimes I^{\otimes 2}) |\psi_0\rangle = \frac{1}{\sqrt{2^3}} (|000\rangle + |001\rangle + \dots + |111\rangle) |10\rangle \\ &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}^T \otimes \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \end{aligned} \tag{83}$$

The first register represents all possible inputs to function $f(x)$.

Iteration (1): 3. Apply Grover's search algorithm

Apply Grover's algorithm using quantum oracle O and inversion around mean r times. The value of r is chosen randomly in order to avoid algorithm trapping [15]. In our example Grover's algorithm (with $r_1 = 1$) produces the state:

$$\begin{aligned} |\psi_2^{(1)}\rangle &= (G \otimes I^{\otimes 2}) \frac{1}{\sqrt{2^3}} (|000\rangle + |001\rangle + \dots + |111\rangle) |10\rangle \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \end{bmatrix}^T \otimes \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \end{aligned} \tag{84}$$

The state after measurement of the first register will either collapse to $|100\rangle |10\rangle$ or $|000\rangle |10\rangle$ with 0.5 probability for both outcomes. In our example assume that the measurement resulted in the state $|000\rangle |10\rangle$.

Iteration (1): 4. Set new threshold

Set $x_2 = 000$ and $y_2 = f(000) = 1$. Change the second register to $|01\rangle$ corresponding to a digit 1. Go to the next iteration.

Convergence

The convergence for that example is achieved after the second iteration. In the second iteration Grover's algorithm (with $r_2 = 2$) produces the state (amplitudes rounded to third digit):

$$|\psi_2^{(2)}\rangle = \begin{bmatrix} 0.972 & -0.088 & -0.088 & -0.088 & -0.088 & -0.088 & -0.088 & -0.088 \end{bmatrix}^T \otimes \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \tag{85}$$

and the global minimum $x = 100$ is measured with probability of $0.972^2 = 0.945$. Advice on a number of iterations is presented in the general algorithm description.

3. Algorithm

Now, we are going to present the generalized version of the algorithm. As an input the algorithm requires:

- Objective function $f(x) : \{0, 1\}^n \rightarrow R$,
- Quantum oracle O acting on n input qubits in the first register.

Theoretically, a searched global optimum x_{min} is measured with high probability at the end. The quantum circuit for that algorithm is presented in Fig. (9).

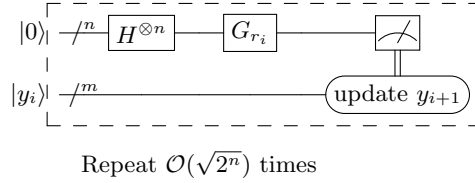


FIG. 9: Quantum circuit of quantum minimization algorithm.

Set starting point

Randomly select input x_1 and set $y_1 = f(x_1)$ as a starting point of algorithm.

Iteration (i): 1. Initialize

Initialize the first register with n input qubits and the second register with m qubits to store threshold y_i .

Iteration (i): 2. Apply Hadamard gates

Apply Hadamard gates on the first register in order to put the first register into a superposition of all possible states representing $f(x)$ inputs.

Iteration (i): 3. Apply Grover's algorithm

Apply Grover's algorithm using quantum oracle O and inversion around mean r_i times. This is denoted in quantum circuit as G_{r_i} gate. The measurement in Grover's algorithm is presented separately on quantum circuit. As mentioned earlier the value of r_i is chosen randomly in order to avoid algorithm trapping. Denote output of Grover's algorithm by $|x\rangle$, convert that state to input x and calculate value of function $y = f(x)$.

Iteration (i): 4. Set new threshold

If $y < y_i$ then set $x_{i+1} = x$ and $y_{i+1} = y$. In other case set value of x and y to previous values $x_{i+1} = x_i$ and $y_{i+1} = y_i$. This comparison operation can be executed on a classical computer.

Convergence

The convergence is achieved in order $\sqrt{2^n}$ iterations. The exact number of iterations is considered in Refs. [13, 14]. Measurement of the first register gives a global minimum x_{min} of the function with high probability.

4. Summary

The quantum parallelism is achieved due to Grover's search, which simultaneously applies gates on all possible inputs in a superposition. The same fact allows this quantum algorithm to find global optimum as it is always searching through all possible inputs. The time complexity depends on a number of iterations in the main loop (around $\sqrt{2^n}$) and a number of inner loops r_i in Grover's search algorithm. It turns out that for the purpose of minimization the numbers r_i are substantially smaller than time complexity $\mathcal{O}(\sqrt{2^n})$ in standalone Grover's search algorithm [15]. Thus, for very large n only a number of iterations in the main loop is significant leading to overall $\mathcal{O}(\sqrt{2^n})$ time complexity of quantum minimization algorithm. This is a quadratic speed-up in comparison to classical optimization methods. The limitations of quantum minimization algorithm are the same as Grover's algorithm, the

decoherence of quantum system can threaten the effectiveness of quantum algorithm destroying its quadratic speedup. As an alternative, one can consider a quantum approximate optimization algorithm (QAOA) [16] designed for noisy computers and used in training of quantum Boltzmann machine neural networks [17].

C. Quantum Fourier transform algorithm

1. Preliminaries

The quantum Fourier transform QFT is a quantum analogue of the classical discrete Fourier transform DFT . The DFT is effectively used in digital signal processing, image processing and data compression. The DFT acts on a vector $x = (x_1, x_2, \dots, x_{N-1})^T$ and maps it to a new vector $y = (y_1, y_2, \dots, y_{N-1})^T$ in a following way:

$$y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j e^{2\pi i \frac{kj}{N}} \quad (86)$$

In order to understand DFT assume input signal consisting of true signal with frequency $f = 10$ and amplitude $A = 2$ described by equation:

$$x_j^{(1)} = 2 \sin(2\pi 10j) \quad (87)$$

and a noise with high frequency $f = 50$ and small amplitude $A = 0.3$ described by equation

$$x_j^{(2)} = 0.3 \sin(2\pi 50j) \quad (88)$$

The true signal and noise signal for $N = 1000$ samples are presented in Fig. (10) and Fig. (11), respectively:

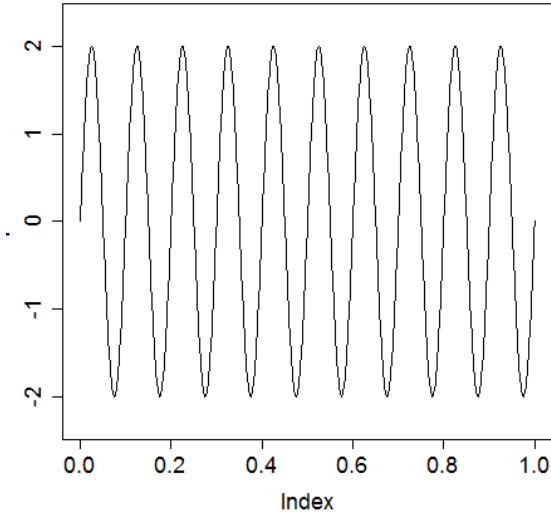


FIG. 10: Noise signal with $f = 10$ and $A = 2$

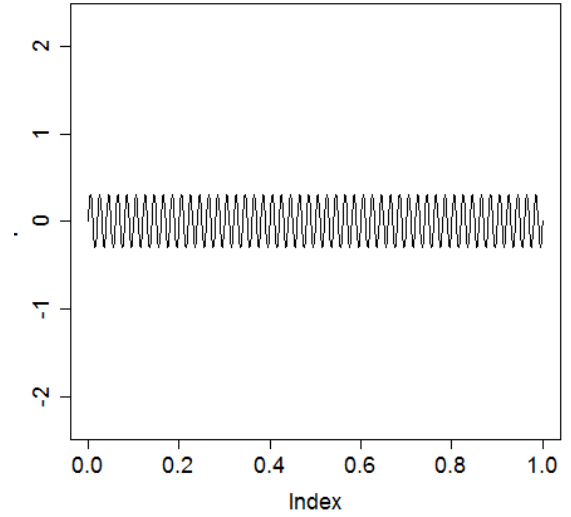
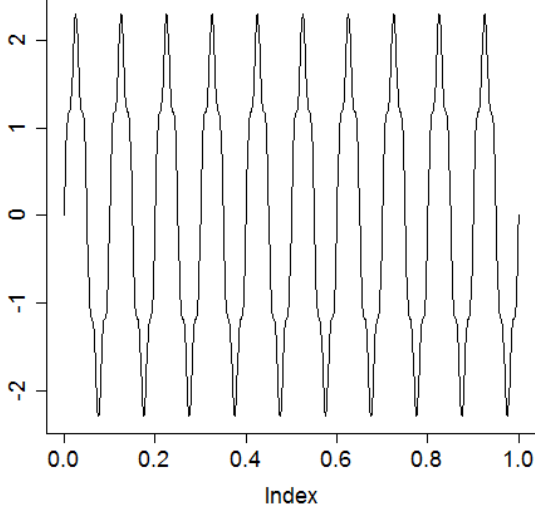
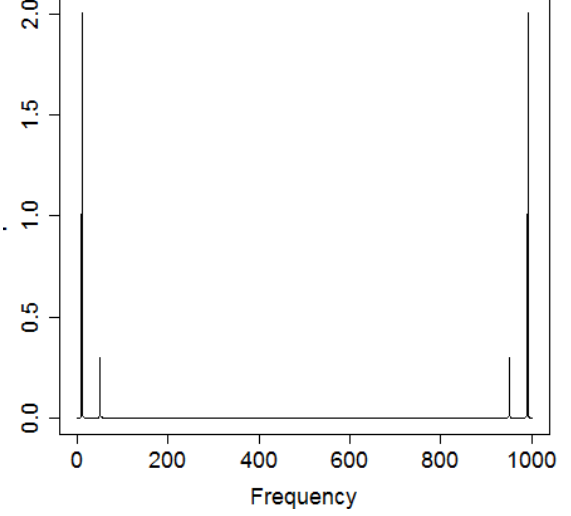


FIG. 11: Noise signal with $f = 50$ and $A = 0.3$

The overall input signal x_j is illustrated in Fig. (12) together with input signal in frequency domain (after application of DFT) in Fig. (13):

FIG. 12: Input signal x_j FIG. 13: Input signal in frequency domain y_k

We can see that exponentials $e^{2\pi i \frac{kj}{N}}$ work as frequency filters for different k leaving peaks located in $k = 10$ and $k = 50$. The mirrored peaks at frequencies 950 and 990 are just the result of using complex numbers. Now, it is easier to reduce noise by discarding peak at frequency 50 (just set the coefficients y_{50} and y_{950} to zero) and apply inverse *DFT*. As a result we get true signal presented in Fig. (10). The quantum version of *DFT* acts on amplitudes of a quantum state $|\psi\rangle = \sum_{j=0}^{N-1} a_j |j\rangle$ and maps them to new amplitudes of a quantum state $|\phi\rangle = \sum_{k=0}^{N-1} b_k |k\rangle$, where

$$b_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} a_j e^{2\pi i \frac{kj}{N}} \quad (89)$$

or equivalently:

$$|j\rangle \rightarrow \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i \frac{kj}{N}} |k\rangle. \quad (90)$$

The arising question is whether we can implement *QFT* on a quantum computer as an effective quantum gate. In other words, we need to find an unitary operator that acts on a quantum state and do the *QFT* operation. It turns out that such operator exists and is defined as:

$$\mathcal{F} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \dots & \omega^{(N-1)(N-1)} \end{bmatrix} \quad (91)$$

where $\omega = e^{\frac{2\pi i}{N}}$.

2. Example

In that example, we are considering the 2-qubit state:

$$|\psi\rangle = a_0 |0\rangle + a_1 |1\rangle + a_2 |2\rangle + a_3 |3\rangle = a_0 |00\rangle + a_1 |01\rangle + a_2 |10\rangle + a_3 |11\rangle . \quad (92)$$

The algorithm finds the Fourier transform:

$$\mathcal{F} |\psi\rangle = |\phi\rangle = \sum_{k=0}^3 b_k |k\rangle \quad (93)$$

with:

$$\begin{aligned} b_0 &= \frac{1}{2}(a_0 + a_1 + a_2 + a_3) \\ b_1 &= \frac{1}{2}(a_0 + a_1 e^{\frac{i\pi}{2}} + a_2 e^{i\pi} + a_3 e^{\frac{3i\pi}{2}}) \\ b_2 &= \frac{1}{2}(a_0 + a_1 e^{i\pi} + a_2 e^{2i\pi} + a_3 e^{3i\pi}) \\ b_3 &= \frac{1}{2}(a_0 + a_1 e^{\frac{3i\pi}{2}} + a_2 e^{3i\pi} + a_3 e^{\frac{9i\pi}{2}}) . \end{aligned} \quad (94)$$

The detailed steps of the algorithm for that particular example are explained below.

Apply Quantum Fourier Transform

The 2-qubit QFT gate can be represented by N -by- N matrix where $N = 2^2 = 4$:

$$\mathcal{F} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \omega & \omega^2 & \omega^3 \\ 1 & \omega^2 & \omega^4 & \omega^6 \\ 1 & \omega^3 & \omega^6 & \omega^9 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \quad (95)$$

as $\omega = e^{\frac{\pi i}{2}} = i$, $\omega^2 = e^{\pi i} = -1$ and $\omega^4 = e^{2\pi i} = 1$ from Euler formula. By applying this gate on input state we get:

$$\mathcal{F} |\psi\rangle = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_0 + a_1 + a_2 + a_3 \\ a_0 + ia_1 - a_2 - ia_3 \\ a_0 - a_1 + a_2 - a_3 \\ a_0 - ia_1 - a_2 + ia_3 \end{bmatrix} . \quad (96)$$

By careful comparison between this matrix and $|\phi\rangle$ state expressed by Eq. (93) we conclude that the desired result has been produced.

Quantum circuit

The QFT gate can be represented by a set of simpler quantum gates. In this example, application of Hadamard gate H on the first qubit, then controlled phase gate $R_{\frac{\pi}{2}}$ also on the first qubit (which is phase shift gate on the first qubit providing that the second qubit is in state $|1\rangle$) and lastly Hadamard gate on the second qubit reproduce the

QFT gate:

$$\begin{aligned}
\mathcal{F} &= (H \otimes I)(C-R_{\frac{\pi}{2}})(I \otimes H)SWAP \\
&= \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & i \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}.
\end{aligned} \tag{97}$$

If we denote the first qubit by x_1 and the second qubit by x_2 , then the quantum circuit representing matrix operations is presented in Fig. (14).

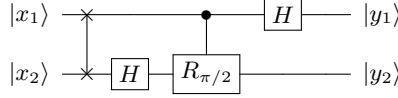


FIG. 14: Quantum circuit of quantum Fourier transform algorithm presented in example.

3. Algorithm

An input to the algorithm is an arbitrary n -qubit state $|\psi\rangle = \sum_{j=0}^{N-1} a_j |j\rangle$ for which the Fourier transform is computed:

$$\mathcal{F} |\psi\rangle = |\phi\rangle = \sum_{k=0}^{N-1} b_k |k\rangle \tag{98}$$

with:

$$b_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} a_j e^{2\pi i \frac{kj}{N}}. \tag{99}$$

Apply Quantum Fourier Transform

Apply QFT gate:

$$\mathcal{F} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \dots & \omega^{(N-1)(N-1)} \end{bmatrix} \tag{100}$$

where $\omega = e^{\frac{2\pi i}{N}}$.

Quantum circuit

The state $|\psi\rangle$ consisting of n qubits can be represented by:

$$|\psi\rangle = |x_1, x_2, \dots, x_n\rangle \tag{101}$$

with x_i denoting state of i -th qubit. Similarly the n -qubit output state $|\phi\rangle$ can be represented by:

$$|\phi\rangle = |y_1, y_2, \dots, y_n\rangle. \quad (102)$$

The input and output states $|\psi\rangle$ and $|\phi\rangle$ will be in a superposition, so that the qubits can be also entangled. The QFT can be applied by a set of basic quantum gates: Hadamard gates and controlled phase shift gates. The reversed order of the input qubits represents multiple $SWAP$ gates applied at the beginning of the quantum circuit:

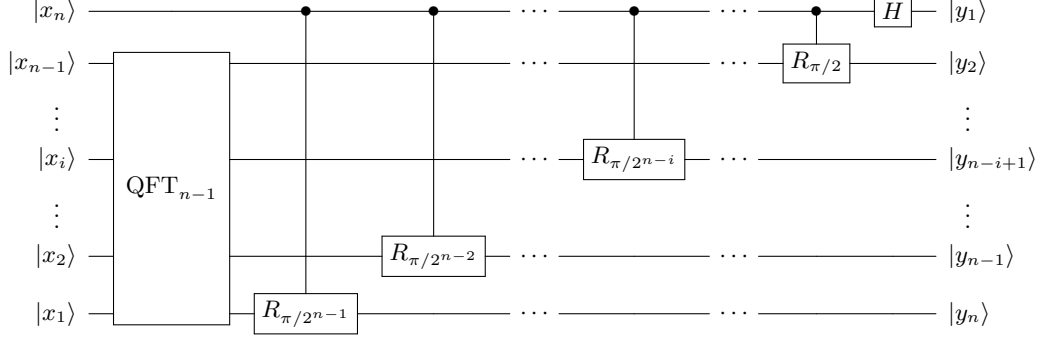


FIG. 15: Quantum circuit of quantum Fourier transform algorithm.

4. Summary

The quantum parallelism for that algorithm results from application of quantum gates acting on superposition (possibly entangled) of input quantum state. Moreover, the quantum inputs state can encode 2^n bits of classical data in just n qubits. Analyzing the quantum circuit it can be concluded that the number of gates required for implementation of QFT algorithm (omitting $SWAP$ gates at the beginning) is equal to $1 + 2 + \dots + n = \frac{n(n+1)}{2}$, which corresponds to $\mathcal{O}(n^2)$ time complexity. The best algorithms improve that complexity to $\mathcal{O}(n \log n)$ [18]. This is an exponential speed-up in comparison to $\mathcal{O}(n2^n)$ time complexity for classical discrete Fourier transform.

D. Quantum phase estimation algorithm

1. Preliminaries

A common problem in applied mathematics is to find eigenvalue of a matrix given its eigenvector. In quantum computation this problem is rephrased into finding an eigenvalue λ of eigenvector $|\phi\rangle$ satisfying $U|\phi\rangle = \lambda|\phi\rangle$ where U is a unitary quantum gate. Due to the fact that U is unitary, eigenvalue can be expressed as $\lambda = e^{2\pi i\theta} = \cos(2\pi\theta) + i\sin(2\pi\theta)$, with phase $0 \leq \theta < 1$. The problem boils down to finding a phase θ with a particular level of precision.

2. Example

We have a unitary matrix:

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (103)$$

and the corresponding eigenvector:

$$|\phi\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (104)$$

We know that the eigenvalue for that problem is $\lambda = 1$, which is equivalent to $\theta = 0$. We apply quantum phase estimation algorithm to get the same result. The detailed steps of the algorithm for that particular example are explained below.

Initialization

We begin with initializing $n = 1$ control qubit that will store a result:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (105)$$

and $m = 1$ qubit encoding eigenvector as:

$$|\phi\rangle = |0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (106)$$

Create superposition

Then, we apply Hadamard gate on control qubit resulting in the state:

$$\begin{aligned} (H \otimes I) |0\rangle |0\rangle &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) |0\rangle. \end{aligned} \quad (107)$$

Apply controlled unitary gates

We would like to apply controlled unitary gate $C-U$, which applies gate U on the eigenvector state only if control qubit is in state $|1\rangle$:

$$C-U \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) |0\rangle = \frac{1}{\sqrt{2}} (|0\rangle |0\rangle + |1\rangle U |0\rangle) = \frac{1}{\sqrt{2}} (|0\rangle + e^{2\pi i \theta} |1\rangle) |0\rangle. \quad (108)$$

This can be rewritten as a following sum (the number 2 is kept in the denominator of exponential, because it would turn out to fit some pattern later):

$$\frac{1}{\sqrt{2}} \sum_{k=0}^1 e^{2\pi i \frac{k2\theta}{2}} |k\rangle |0\rangle. \quad (109)$$

Now, a result of the control qubit measurement does not depend on phase θ because the amplitude in exponential $e^{2\pi i k \frac{2\theta}{2}}$ is killed by its complex conjugate $e^{-2\pi i k \frac{2\theta}{2}}$ during calculation of probability. If only we can express the control qubit in a different basis which will be dependent on phase θ . A closer look unveils that the control qubit state has a form of Fourier-transformed state:

$$|j\rangle = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} e^{2\pi i \frac{kj}{N}} |k\rangle \quad (110)$$

with $N = 2$ and $j = 2\theta$. The last step is to recover $|j\rangle = |2\theta\rangle$ by applying inverse operation to Fourier transform.

Apply inverse quantum Fourier transform gate

The inverse quantum Fourier transform creates a different basis:

$$\mathcal{F}^{-1} |k\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} e^{-2\pi i \frac{kj}{N}} |j\rangle. \quad (111)$$

For our example the inverse quantum Fourier transform gate for 2 qubits is expressed as:

$$\mathcal{F}^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & \omega^{-1} \end{bmatrix} \quad (112)$$

where $\omega = e^{\frac{2\pi i}{2}} = e^{\pi i}$. Thus:

$$\mathcal{F}^{-1} \frac{1}{\sqrt{2}} \sum_{k=0}^1 e^{2\pi i \frac{k2\theta}{2}} |k\rangle |0\rangle = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & e^{-\pi i} \end{bmatrix} \begin{bmatrix} 1 \\ e^{2\pi i \theta} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 + e^{2\pi i \theta} \\ 1 + e^{\pi i} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (113)$$

which is equivalent to:

$$\frac{1}{2} \sum_{j=0}^1 \sum_{k=0}^1 e^{2\pi i \frac{k(2\theta-j)}{2}} |j\rangle |0\rangle. \quad (114)$$

Measurement

Measurement of the control qubit collapses the state of system to $|2\theta\rangle |0\rangle$. Note that the probability of measuring control qubit to be in the state $|2\theta\rangle$ is calculated as:

$$\begin{aligned} P(|2\theta\rangle) &= \left| \frac{1}{2} \langle 2\theta | \sum_{j=0}^1 \sum_{k=0}^1 e^{2\pi i \frac{k(2\theta-j)}{2}} |j\rangle \right|^2 \\ &= \left| \frac{1}{2} \sum_{k=0}^1 e^0 \langle 2\theta | 2\theta \rangle \right|^2 \\ &= \left| \frac{1+1}{2} \right|^2 \\ &= 1. \end{aligned} \quad (115)$$

It means that measuring the control qubit always gives a correct θ . In our example the qubit will be measured in $|0\rangle$ state as $\theta = 0$. In cases where θ is not an integer or we operate on larger eigenvectors we need to increase a number of control qubits n as well as a number of controlled quantum gates $C-U$. General algorithm is presented in the next section.

3. Algorithm

The general algorithm is represented by the quantum circuit in Fig. (16). Similarly to the example presented in the previous subsection, we would require a quantum gate U and its eigenvector $|\phi\rangle$ as an input. The algorithm reveals a phase θ allowing for calculation of an eigenvalue $\lambda = e^{2\pi i\theta}$.

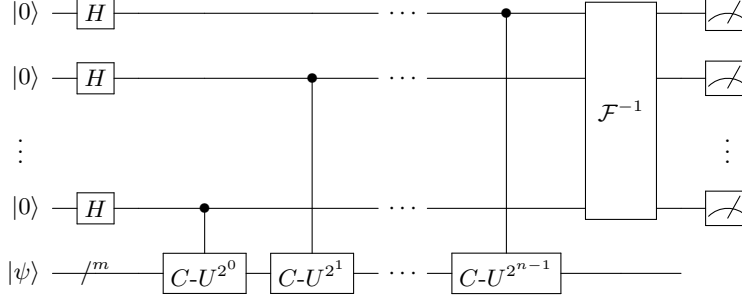


FIG. 16: Quantum circuit of quantum phase estimation algorithm.

Initialization

Initialize n control qubits $|0\rangle$ and m qubits encoding eigenvector $|\phi\rangle$. The initial state of the system is $\psi_0 = |0\rangle^{\otimes n} |\phi\rangle$.

Create superposition

Apply n -bit Hadamard gate $H^{\otimes n}$ on n control qubits resulting in the state:

$$|\psi_1\rangle = \frac{1}{\sqrt{2^n}}(|0\rangle + |1\rangle)^{\otimes n} |\phi\rangle \quad (116)$$

Apply controlled unitary gates

Controlled unitary gate $C-U$ applies gate U on the state $|\phi\rangle$ only if the control qubit is in the state $|1\rangle$. Having n control qubits we apply n control unitary gates $C-U^{2^j}$ with $0 \leq j \leq n-1$ on the eigenstate $|\phi\rangle$. Noticing that:

$$U^{2^j} |\phi\rangle = U^{2^j-1} U |\phi\rangle = U^{2^j-1} e^{2\pi i\theta} |\phi\rangle = e^{2\pi i2^j\theta} |\phi\rangle \quad (117)$$

and moving the phases from eigenvector to control qubits we eventually end up with the state in a superposition:

$$|\psi_2\rangle = \frac{1}{\sqrt{2^n}} \underbrace{(|0\rangle + e^{2\pi i2^{n-1}\theta} |1\rangle)}_{\text{qubit 1}} \otimes \dots \otimes \underbrace{(|0\rangle + e^{2\pi i2^1\theta} |1\rangle)}_{\text{qubit n-1}} \otimes \underbrace{(|0\rangle + e^{2\pi i2^0\theta} |1\rangle)}_{\text{qubit n}} \otimes |\phi\rangle. \quad (118)$$

The Eq. (118) can be simplified by rewriting all 2^n n -qubits states $\{|00\dots 0\rangle, |00\dots 1\rangle, \dots, |11\dots 1\rangle\}$ to states enumerated with an integer $|k\rangle$ made up from summing 2^j 's in exponentials. As an example $e^{2\pi i0\theta} |00\dots 0\rangle$ is rewritten as $e^{2\pi i0\theta} |0\rangle$, $e^{2\pi i2^0\theta} |0\dots 01\rangle$ is rewritten as $e^{2\pi i1\theta} |1\rangle$ and $e^{2\pi i(2^1+2^0)\theta} |0\dots 11\rangle$ is rewritten as $e^{2\pi i3\theta} |3\rangle$. It can be generalized as:

$$|\psi_2\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{2\pi ik\theta} |k\rangle |\phi\rangle = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} e^{2\pi i \frac{k2^n\theta}{2^n}} |k\rangle |\phi\rangle. \quad (119)$$

Apply inverse quantum Fourier transform gate

Control qubits in the state $|\psi_2\rangle$ have form of a Fourier-transformed state. Applying inverse quantum Fourier transform gate QFT^{-1} maps to a new basis where state containing phase gathers all amplitudes. The QFT^{-1} gate applied on the previous state yields:

$$\begin{aligned} |\psi_3\rangle &= \frac{1}{2^n} \sum_{j=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{2\pi ik\theta} e^{-2\pi i \frac{kj}{2^n}} |j\rangle |\phi\rangle \\ &= \frac{1}{2^n} \sum_{j=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{2\pi i \frac{k(2^n\theta-j)}{2^n}} |j\rangle |\phi\rangle. \end{aligned} \quad (120)$$

Phase approximation

Due to the fact that states $|j\rangle$ can encode only discrete set of variables, whereas θ is a continuous variable in range $0 \leq \theta < 1$ we approximate $2^n\theta = a + 2^n\delta$ where a is the nearest integer of $2^n\theta$ and δ is rounding error with $0 \leq |2^n\delta| \leq \frac{1}{2}$. We can rewrite the previous state as:

$$|\psi_3\rangle = \frac{1}{2^n} \sum_{j=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{2\pi i \frac{k(a-j)}{2^n} + 2\pi i k \delta} |j\rangle |\phi\rangle. \quad (121)$$

Measurement

Now, measuring control qubits yields state $|a\rangle$ with probability (for simplicity we assume that $\delta = 0$ i.e. there is no rounding error):

$$\begin{aligned} P(|a\rangle) &= \left| \frac{1}{2^n} \langle a | \sum_{j=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{2\pi i \frac{k(a-j)}{2^n}} |j\rangle \right|^2 \\ &= \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} e^{2\pi i \frac{k(a-a)}{2^n}} \langle a | a \rangle \right|^2 \\ &= \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} e^0 \right|^2 \\ &= 1. \end{aligned} \quad (122)$$

It means that measuring control qubits always leaves them in the state $|a\rangle = |2^n\theta\rangle$ from which we can read off the phase and then calculate an eigenvalue. The state of the whole system after measurement is $|2^n\theta\rangle |\phi\rangle$. In case rounding error is not zero the state $|a\rangle$ will be yielded with probability:

$$\begin{aligned} P(|a\rangle) &= \left| \frac{1}{2^n} \langle a | \sum_{j=0}^{2^n-1} \sum_{k=0}^{2^n-1} e^{2\pi i \frac{k(a-j)}{2^n} + 2\pi i k \delta} |j\rangle \right|^2 \\ &= \left| \frac{1}{2^n} \sum_{k=0}^{2^n-1} e^{2\pi i k \delta} \right|^2 \\ &= \frac{1}{2^{2n}} \left| \frac{1 - e^{2\pi i 2^n \delta}}{1 - e^{2\pi i \delta}} \right|^2. \end{aligned} \quad (123)$$

It can be shown [19] that $P(|a\rangle) \geq \frac{4}{\pi^2} \approx 0.405$ and converges to 1 as number of control qubits n increases.

4. Summary

The quantum phase estimation algorithm allows us to calculate the eigenvalue for a given eigenvector of a quantum gate. It is used in quantum Principal Components Analysis algorithm described in section VD. Assuming the C - U gates are already prepared, the time complexity depends largely on the time complexity of quantum Fourier transform which is $\mathcal{O}(n \log n)$ for best-known algorithms [18].

V. QUANTUM MACHINE LEARNING ALGORITHMS

This section provides an overview of selected unsupervised and supervised quantum machine learning algorithms, however there is also a method of scores extraction for quantum PCA algorithm proposed as well as a new cost function in feed-forward quantum neural networks is introduced. The unsupervised machine learning algorithms are

quantum versions of k -means and k -median algorithms as well as quantum principal component analysis (qPCA). The supervised algorithms presented are quantum support vector machines (qSVM) and quantum neural networks (qNN). The quantum machine learning algorithms utilize the quantum algorithms described in the previous section as subroutines. The following names of subroutines have been assigned:

- *GroverOptim*: Quantum minimization algorithm,
- *PhaseEstim*: Quantum phase estimation algorithm.

Additional quantum subroutines have been introduced purely for the purpose of quantum machine learning algorithms:

- *SwapTest*,
- *DistCalc*,
- *MedianCalc*.

These are carefully explained in the following subsections. The mapping between quantum machine learning algorithms and quantum subroutines is presented (U - Used, O - Optional):

TABLE III: Quantum machine learning algorithms and quantum subroutines mapping

	<i>GroverOptim</i>	<i>PhaseEstim</i>	<i>SwapTest</i>	<i>DistCalc</i>	<i>MedianCalc</i>
q k -means	U		U	U	
q k -medians	U		U	U	U
qSVM	U				
qPCA		U	O		
qNN					

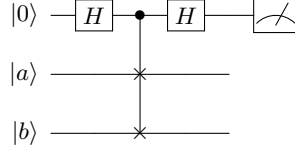
A. Quantum k -means algorithm

1. Preliminaries

Classical k -means algorithm belongs to the family of unsupervised machine learning algorithms. It aims to classify data to k clusters based on an unlabeled set of training vectors. The training vectors are reassigned to the nearest centroid in each iteration and then a new centroid is calculated averaging the vectors belonging to the current cluster. The time complexity $\mathcal{O}(NM)$ of the classical algorithm using Lloyd's version of the algorithm is linearly dependent on a number of features N and a number of training examples M [20]. The most resource consuming operation for k -means algorithm is calculation of a distance between vectors and we are expecting to gain a speed-up by retrieving the distance using a quantum computer. The quantum k -means algorithm turns out to provide an exponential speed-up for very large dimension of a training vector. The algorithm is based on two new quantum subroutines calculating a distance between vectors: *SwapTest* and *DistCalc* as well as the quantum subroutine *GroverOptim* already described in section IV B and assigning vector to the closest centroid of cluster.

2. Quantum subroutine: *SwapTest*

The subroutine *SwapTest* is a simple quantum routine expressing overlap of two states $\langle a|b \rangle$ in terms of measurement probability of control qubit being in state $|0\rangle$. The overlap is a measure of similarity between quantum states and it will be used in calculation of a distance between classical vectors in *DistCalc* subroutine. The algorithm has been firstly used in Ref. [21]. The quantum circuit of the algorithm is presented in Fig. (17).

FIG. 17: Quantum circuit of *SwapTest* quantum subroutine.

The steps of the subroutine are presented below.

Initialize

Initialize two states $|a\rangle$ and $|b\rangle$ as well as a control qubit $|0\rangle$ resulting in the state:

$$|\psi_0\rangle = |0, a, b\rangle. \quad (124)$$

The states $|a\rangle$ and $|b\rangle$ consist of n qubits each.

Apply Hadamard gate

Apply Hadamard gate on the control qubit resulting in a superposition:

$$|\psi_1\rangle = (H \otimes I^{\otimes n} \otimes I^{\otimes n}) |\psi_0\rangle = \frac{1}{\sqrt{2}} (|0, a, b\rangle + |1, a, b\rangle). \quad (125)$$

Apply SWAP gate

Apply controlled *SWAP* gate on $|a\rangle$ and $|b\rangle$ states which swaps a and b providing that the control qubit is in state $|1\rangle$. As a result:

$$|\psi_2\rangle = \frac{1}{\sqrt{2}} (|0, a, b\rangle + |1, b, a\rangle). \quad (126)$$

Apply Hadamard gate

Apply second Hadamard gate on the control qubit resulting in the state:

$$|\psi_3\rangle = \frac{1}{2} |0\rangle (|a, b\rangle + |b, a\rangle) + \frac{1}{2} |1\rangle (|a, b\rangle - |b, a\rangle). \quad (127)$$

Measurement

Measure the control qubit. The probability of measuring control qubit being in state $|0\rangle$ is given by:

$$\begin{aligned} P(|0\rangle) &= \left| \frac{1}{2} \langle 0|0\rangle (|a, b\rangle + |b, a\rangle) + \frac{1}{2} \langle 0|1\rangle (|a, b\rangle - |b, a\rangle) \right|^2 \\ &= \frac{1}{4} |(|a, b\rangle + |b, a\rangle)|^2 \\ &= \frac{1}{4} (\langle b|b\rangle \langle a|a\rangle + \langle b|a\rangle \langle a|b\rangle + \langle a|b\rangle \langle b|a\rangle + \langle a|a\rangle \langle b|b\rangle) \\ &= \frac{1}{2} + \frac{1}{2} |\langle a|b\rangle|^2 \end{aligned} \quad (128)$$

Thus, we successfully linked an overlap $\langle a|b\rangle$ with measurement probability of the control qubit in the final quantum state. The probability $P(|0\rangle) = 0.5$ means that the states $|a\rangle$ and $|b\rangle$ are orthogonal, whereas the probability $P(|0\rangle) = 1$ indicates that the states are identical. The subroutine should be repeated several times to obtain a good estimator of probability. The advantage of using swap test is that the states $|a\rangle$ and $|b\rangle$ can be unknown before procedure and simple measurement is performed on the control qubit which has two eigenstates. The time complexity is negligible as the procedure does not depend on a number of qubits representing the input states, however we may pay attention to the preparation time of identical copies of $|a\rangle$ and $|b\rangle$.

3. Quantum subroutine: *DistCalc*

With subroutine *SwapTest* presented we can move on to the algorithm retrieving the Euclidean distance $|a - b|^2$ between two real valued vectors a and b . The algorithm was described by Lloyd, Mohseni and Rebentrost [22].

Representation of classical data as quantum states

The classical information in vector a is encoded as [23, 24]:

$$|a|^{-1}a \rightarrow |a\rangle = \sum_{i=1}^N |a|^{-1}a_i |i\rangle. \quad (129)$$

Norm of quantum state is normalized with this definition $\langle a|a\rangle = |a|^{-2}a^2 = 1$, which leads to the correct definition of a quantum state. The N dimensional training vector can be translated into $n = \log_2 N$ qubits. As an example the vector with eight features can be stored in three qubits containing $2 * 2 * 2 = 8$ entries to express vector components a_i as probability amplitudes.

Initialize

Initialize two quantum states:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|0, a\rangle + |1, b\rangle), \quad (130)$$

$$|\phi\rangle = \frac{1}{\sqrt{Z}}(|a| |0\rangle + |b| |1\rangle) \quad (131)$$

with $Z = |a|^2 + |b|^2$.

Use quantum subroutine *SwapTest*

Evaluate an overlap $\langle\phi|\psi\rangle$ using subroutine *SwapTest*.

Calculate distance

Calculate Euclidean distance by noting that:

$$|a - b|^2 = 2Z |\langle\phi|\psi\rangle|^2. \quad (132)$$

This holds true, because:

$$\langle\phi|\psi\rangle = \frac{1}{\sqrt{2Z}}(|a| |a\rangle - |b| |b\rangle) \quad (133)$$

and using the fact how the classical information has been prepared in Eq. (129) the inner product could be expressed as:

$$\langle\phi|\psi\rangle = \frac{1}{\sqrt{2Z}}(a - b) \quad (134)$$

so that $|\langle\phi|\psi\rangle|^2 = \frac{1}{2Z}|a - b|^2$. Moreover, using this algorithm it is also easy to calculate the inner product between two vectors noticing that:

$$a^T b = \frac{1}{2}(|a|^2 + |b|^2 - |a - b|^2). \quad (135)$$

To summarize, in *DistCalc* subroutine we prepare two states, apply subroutine *SwapTest* and repeat that procedure to obtain an acceptable estimate of probability. Providing that the states are already prepared the subroutine *SwapTest* does not depend on the size of a feature vector. The preparation of states in subroutine *DistCalc* is proved to have $\mathcal{O}(\log N)$ time complexity [22], which makes sense as the classical information is encoded in $n = \log_2 N$ qubits and we expect the time complexity to be proportional. The classical algorithms require $\mathcal{O}(N)$ to calculate Euclidean distance between two vectors, thus there is an exponential speed-up. The measurement during swap test is causing the decoherence of input states, thus the quantum memory should contain multiple copies of input states, however it does not change time complexity for large N , as the number of states preparation will be always much smaller than N .

4. Algorithm

Now, having two quantum subroutines *SwapTest* and *DistCalc* required to calculate distance and quantum optimization subroutine *GroverOptim* to choose the closest centroid of cluster we are ready to put down the steps of quantum k -means algorithm.

Initialize

Choose number of k clusters and randomly initialize k cluster centroids $\mu_1, \mu_2, \dots, \mu_k \in R^N$. These initializations can be done by any methods used in classical k -means algorithm.

Main Loop (s): Do until convergence is reached

Inner Loop (i): Choose closest cluster centroid

Loop over training examples $i = \{1, \dots, M\}$ and for each training example $x^{(i)}$ use quantum subroutine *DistCalc* to calculate k distances $\|x^{(i)} - \mu_k\|$ to each cluster centroids and then use quantum subroutine *GroverOptim* to choose index $c^{(i)} = \{1, \dots, k\}$ of cluster centroid that minimize the distance between training example and cluster centroid:

$$c^{(i)} := \arg \left\{ \min_k \|x^{(i)} - \mu_k\|^2 \right\}. \quad (136)$$

Inner Loop (j): Calculate new cluster centroids

Loop over clusters $j = \{1, \dots, k\}$ and calculate mean μ_j of points assigned to cluster j . The calculated μ_j becomes new cluster centroid. Specifically, for each centroid j we set:

$$\mu_j = \frac{1}{|C_j|} \sum_{i \in C_j} x^{(i)} \quad (137)$$

where $|C_j|$ is the number of training vectors belonging to cluster j and C_j is the set of these vectors. For example, imagine that there are four training examples belonging to cluster $j = 2$: $C_2 = \{x^{(2)}, x^{(6)}, x^{(10)}, x^{(11)}\}$. The new cluster is calculated as $\mu_2 = \frac{1}{4}(x^{(2)} + x^{(6)} + x^{(10)} + x^{(11)})$. This operation is executed on a classical computer.

Convergence:

Convergence is reached when subsequent iterations of algorithm does not change the location of cluster centroids. Mathematically speaking, the algorithm converges when for each cluster $j = \{1, \dots, k\}$ the distance between cluster centroids from iteration $s + 1$ and the previous iteration s is lower than some arbitrary threshold η :

$$\|\mu_j^{s+1} - \mu_j^s\| < \eta \quad (138)$$

5. Summary

The quantum k -means algorithm provides an exponential speed-up in comparison to classical k -means algorithm. The Lloyd's version of k -means algorithm executed on a classical computer has time complexity $\mathcal{O}(MNk)$. This just corresponds to the fact that for each training example we need to calculate distances between N dimensional feature vectors and k clusters. The number of algorithm iterations is assumed to be negligible in case we deal with large M , N and k (of course the algorithm can do many iterations before it converges, but we assume that it is not the main source of time complexity). The quantum parallelism is achieved thanks to the clever calculation of distances between large dimensional vectors. Thus, the time complexity of the quantum algorithm is $\mathcal{O}(Mk \log(N))$, which for very large N is an exponential speed-up. Note that in the following time complexities the minimization procedure is not taken into account, so for a large number of clusters quantum minimization subroutine provides even more speed-up providing that the quantum oracles for Grover's algorithm are prepared efficiently. However, for a small number of clusters it could be advisable to use classical optimization algorithms, as Dürr and Høyer minimization algorithm does it best for a large number of objective function inputs. We should also remember about limitations of quantum minimization algorithm in the presence of quantum noise. Further improvements of the algorithm exist, which uses a quantum adiabatic algorithm and quantum output [22].

B. Quantum k -medians algorithm

1. Preliminaries

The k -medians algorithm is similar to k -means algorithm, however instead of setting new cluster centroids by calculating mean, a median is evaluated. The k -means algorithm has a disadvantage that in some cases, the calculated mean may lie outside the desired region, whereas cluster centroid calculated as median always belongs to a training vectors set. It has even further implications for quantum algorithms, as the calculation of mean is executed on a classical computer, whereas the calculation of median can be implemented using *MedianCalc* subroutine [25]. The quantum version of k -medians algorithm requires the same quantum subroutines as quantum k -means algorithm: *SwapTest*, *DistCalc* and *GroverOptim* plus an additional *MedianCalc*, which also uses *GroverOptim*. The quantum minimization algorithm *GroverOptim* is optional to use in case of k -means algorithm, whereas for k -medians it may be used to choose closest centroids and must be used to calculate median.

2. Quantum subroutine: *MedianCalc*

The quantum subroutine to calculate median Q from set of N dimensional points $\{a_1, a_2, \dots, a_m\}$ have been introduced in [25]. By definition, the median is a point within the set $\{a_1, a_2, \dots, a_m\}$ whose distance to all other points is minimum. As a consequence, in order to calculate median we must evaluate the following sum for each a_i :

$$S_i = \sum_{j=1}^m \|a_i - a_j\| \quad (139)$$

and take a minimum. The index i_{min} for the smallest S_i denotes a median $Q = a_{i_{min}}$. The inputs to the algorithm are listed below:

- Set of N dimensional points $\{a_1, a_2, \dots, a_m\}$,
- Quantum oracle O used in *GroverOptim* subroutine.

Below, we are giving the description of the subroutine.

Calculation of distances

Use *DistCalc* subroutine to calculate all required distances for sums $\{S_1, \dots, S_m\}$.

Use quantum subroutine *GroverOptim*

Use quantum minimization subroutine to choose smallest S_i .

Median

The index i_{min} for the smallest S_i will define median $Q = a_{i_{min}}$. Generally, the operation of choosing minimal S_i executed on classical computer has time complexity of $\mathcal{O}(m)$. Assuming that a black-box quantum oracle O has been already provided, the *GroverOptim* subroutine can be used as it is a minimization problem. The optimization operation on quantum computer has time complexity $\mathcal{O}(\sqrt{m})$. There is also a gain in calculation of Euclidean distance between vectors using quantum subroutine *DistCalc*. Assuming that dimension of feature vector N is much larger than a number of vectors m we gain an additional exponential speed-up as presented in section V A 3.

3. Algorithm

The algorithm is similar to quantum k -means algorithm.

Initialize

Choose number of k clusters and randomly generate k cluster centroids $\mu_1, \mu_2, \dots, \mu_k \in R^N$ from training examples set $x^{(1)}, x^{(2)}, \dots, x^{(M)}$.

Main Loop (s): Do until convergence is reached

Inner Loop (i): Choose closest cluster centroid

Loop over training examples $i = \{1, \dots, M\}$ and for each training example $x^{(i)}$ use quantum subroutine *DistCalc* to calculate k distances $\|x^{(i)} - \mu_k\|$ to each cluster centroids and then use quantum subroutine *GroverOptim* to choose index $c^{(i)} = \{1, \dots, k\}$ of cluster centroid that minimize the distance between training example and cluster centroid:

$$c^{(i)} := \arg \left\{ \min_k \|x^{(i)} - \mu_k\|^2 \right\}. \quad (140)$$

Inner Loop (j): Calculate new cluster centroids

Loop over clusters $j = \{1, \dots, k\}$ and calculate median μ_j of points assigned to cluster j using quantum subroutine *MedianCalc*.

Convergence:

Convergence is reached when subsequent iterations of algorithm does not change a location of cluster centroids. For k -medians algorithm it just means that each cluster $j = \{1, \dots, k\}$ in iteration $s + 1$ is the same as in the previous iteration s :

$$\mu_j^{s+1} = \mu_j^s. \quad (141)$$

4. Summary

The time complexity of quantum k -median algorithm is similar to quantum k -mean algorithm except for the time complexity difference between calculation of medians and means. In comparison to its classical counterpart, there is an exponential speed up coming from the same source as in quantum k -means algorithm and additional quadratic gain coming from the calculation of median on a quantum computer. The quantum parallelism of *MedianCalc* is achieved by application of Grover's search in quantum minimization subroutine. Although k -median algorithm has the advantage of having clusters always in the desirable region, the calculation of median on a quantum computer can be longer than the simple calculation of mean on the classical computer due to the exhaustive evaluation of distances. This doubt can be disregarded if we assume that the distances are already given to the quantum k -median algorithm. This is a very interesting assumption which is based on the fact that instead of giving a set of training examples to k -median procedure, we can just calculate distances between all training examples and take it as the only input to the algorithm. Thus, the k -median algorithm, in this case, is more attractive than k -means algorithm if we can only obtain distances instead of vectors as a training data.

C. Quantum support vector machines

1. Preliminaries

The support vector machines algorithm (SVM) is one of supervised machine learning algorithms used for linear discrimination problems. The algorithm is based on finding a hyperplane that discriminates between two classes of feature vectors and is used as a decision boundary for future data classification. The SVM is formulated as maximizing the distance between the hyperplane and closest data points called support vectors. The objective function could be convex or non-convex depending on the kernel used in SVM algorithm. The non-convex functions tend to converge to a local optimum, thus the classical SVM balances between optimization effectiveness and the prediction accuracy. Quantum version of support vector machines using minimization quantum subroutine *GroverOptim* and provides convergence to a global optimum for non-convex cost function [26].

2. Algorithm

For simplicity, assume linear support vector machines with hyperplanes defined by:

$$\theta^T x - b = \pm 1 \quad (142)$$

and a distance between two hyperplanes expressed by $\|\frac{2}{\theta}\|$. Note that x and θ are vectors, whereas b is a numeric constant. Minimizing θ leads to a maximum margin and adding constraints would ensure that the margin correctly classifies the data. This can be represented by the following optimization problem:

$$\min_{\theta, b} \frac{1}{2} \|\theta\|^2 \quad (143)$$

subject to a constraint:

$$y^{(i)}(\theta^T x^{(i)} - b) \geq 1 \quad (144)$$

for all training examples $i = \{1, \dots, M\}$ and $y^{(i)} = \{-1, 1\}$. The constraint can be incorporated into an objective function using Lagrange multipliers $\alpha^{(i)}$, which results in following formulation of the problem:

$$\min_{\theta, b} \max_{\alpha^{(i)} \geq 0} \left(\frac{1}{2} \|\theta\|^2 - \sum_{i=1}^M [\alpha^{(i)}(\theta^T x^{(i)} - b) - 1] \right). \quad (145)$$

Note, that the only non-zero $\alpha^{(i)}$ will corresponds to $x^{(i)}$ being support vectors. In order to solve the maximization of objective function F with respect to $\alpha^{(i)}$, we set the following derivatives to zero:

$$\begin{aligned} \frac{\partial F}{\partial \theta^{(i)}} &= \theta^{(i)} - \alpha^{(i)} y^{(i)} x^{(i)} = 0 \\ \frac{\partial F}{\partial b} &= \sum_{i=1}^M \alpha^{(i)} y^{(i)} = 0. \end{aligned} \quad (146)$$

As a consequence, we can express weights as:

$$\theta = \sum_{i=1}^M \alpha^{(i)} y^{(i)} x^{(i)} \quad (147)$$

and the dual problem as:

$$\min_{\alpha^{(i)}} \left\{ \frac{1}{2} \sum_{i,j} \alpha^{(i)} \alpha^{(j)} y^{(i)} y^{(j)} (x^{(i)})^T x^{(j)} - \sum_{i=1}^M \alpha^{(i)} \right\} \quad (148)$$

providing that:

$$\alpha^{(i)} \geq 0 \quad (149)$$

for each training example $i = 1, \dots, M$ and:

$$\sum_{i=1}^M \alpha^{(i)} y^{(i)} = 0. \quad (150)$$

The optimization problem can be generalized to an arbitrary kernel function $K(x^{(i)}, x^{(j)})$ introducing the non-linearity to the problem. The dot product in the previous dual problem is replaced with a kernel function:

$$\min_{\alpha^{(i)}} \left\{ \frac{1}{2} \sum_{i,j} \alpha^{(i)} \alpha^{(j)} y^{(i)} y^{(j)} K(x^{(i)}, x^{(j)}) - \sum_{i=1}^M \alpha^{(i)} \right\}. \quad (151)$$

As an example, we can take Gaussian kernel function expressed as:

$$K(x^{(i)}, x^{(j)}) = \exp \left(-\gamma \|x^{(i)} - x^{(j)}\|^2 \right) \quad (152)$$

so that additional calculation of Euclidean distances is required. Now, we are going to give the detailed description of each algorithm's step.

Initialize the free parameters and kernel function

Initialize the value of any free parameters that are used in a kernel function. Choose an appropriate kernel function for the problem and calculate a kernel matrix.

Representation of classical data and parameters as quantum states

This step implies discretization of objective function and writing its components to qubits. The classical information can be represented as binary string:

$$x \rightarrow a = (a_1, a_2, \dots, a_k)^T \quad (153)$$

with $a_i = \{0, 1\}$ for $i = 1 \dots, k$. Then the binary strings can be directly translated into k qubit quantum state:

$$|a_1 a_2 \dots a_k\rangle \quad (154)$$

from a 2^k dimensional Hilbert space spanned by basis $\{|00 \dots 0\rangle, |10 \dots 0\rangle, \dots, |11 \dots 1\rangle\}$.

Scanning objective function space

The quantum minimization subroutine is searching through the objective function space to find an optimal set of $\alpha^{(i)}$, which solves for parameters θ and b . The subroutine *GroverOptim* creates the superposition of all possible inputs to a quantum oracle O representing objective function and finds a global minimum for the SVM optimization problem. The measurement in subroutine *GroverOptim* reveals the solution with high probability.

3. Summary

The quantum subroutine *GroverOptim* decreases the time complexity quadratically from $\mathcal{O}(N)$ to $\mathcal{O}(\sqrt{N})$ and provides a global minimum of the formulated SVM optimization problem. Still, the most time-consuming part of SVM algorithm is the calculation of a kernel matrix, which has the time complexity of $\mathcal{O}(M^2 N)$ on a classical computer. The limitations of the algorithm are the same as for quantum subroutine *GroverOptim*, i.e. the algorithm would not provide satisfying results in the presence of quantum noise. Although, we assume that a quantum oracle representing objective function of SVM is an input to the algorithm, we should not ignore that building such an oracle could be complex or impossible. An unconstrained exponential speed-up is presented in another quantum algorithm for SVM presented in Ref. [27].

D. Quantum principal component analysis

1. Preliminaries

Principal component analysis (PCA) is widely used dimensionality reduction procedure. It takes N dimensional feature vectors (possibly correlated) from a training set, applies an orthonormal transformation and outputs R dimensional data. The compressed data can be further used in other machine learning algorithms allowing to draw the same conclusions as on the full data, however executing algorithms much faster as in some cases $R \ll N$. The PCA relies on the eigendecomposition of the correlation (or covariance) matrix, thus a quantum speed-up should be obtained in that area. The quantum phase estimation subroutine *PhaseEstim* deals with eigenvectors and eigenvalues and indeed is used by quantum principal component analysis (qPCA). The qPCA algorithm is exponentially faster than any known classical algorithm and has been presented by Lloyd, Mohseni and Rebentrost [28]. The algorithm uses the quantum subroutine *PhaseEstim* on an exponent of density matrix ρ and produces the quantum states containing all eigenvectors and eigenvalues of the density matrix ρ . We will see that a density matrix ρ is equivalent to covariance/correlation matrix.

2. Algorithm

Demean and normalization of classical data

Firstly, a set of N dimensional training vectors should be demeaned in order to properly use it in qPCA. For each

training example $x^{(i)}$ with $i = \{1, \dots, M\}$ we subtract the N dimensional vector with means \bar{x} :

$$\begin{aligned} x^{(i)} &\rightarrow x^{(i)} - \bar{x} \\ \bar{x} &= \frac{1}{M} \sum_{i=1}^M x^{(i)}. \end{aligned} \quad (155)$$

Secondly, the data should be normalized to build proper quantum states. We divide each training example $x^{(i)}$ by a norm of vector:

$$x^{(i)} \rightarrow \left| x^{(i)} \right|^{-1} x^{(i)} \quad (156)$$

with the norm defined as:

$$|x| = \sqrt{\sum_{k=1}^N x_k^2}. \quad (157)$$

We can also standardize the data in order to get correlation matrix, but this is optional.

Representation of classical data as quantum states

The components x_k with $k = \{1, \dots, N\}$ of classical training vector x are encoded as amplitudes of a quantum state [23, 24]:

$$x \rightarrow |x\rangle = \sum_{k=1}^N x_k |k\rangle. \quad (158)$$

Due to the previous normalization of classical data, the quantum state is correctly defined with probabilities summing up to 1: $\langle x|x\rangle = x^2 = 1$. The N dimensional training vector can be translated into $n = \log_2 N$ qubits.

Representation of covariance/correlation matrix as density matrix

Having all training examples encoded as states, we are building a mixed state defined by the following density matrix:

$$\rho = \frac{1}{M} \sum_{i=1}^M |x^{(i)}\rangle \langle x^{(i)}|. \quad (159)$$

The tensor product $|x^{(i)}\rangle \langle x^{(i)}|$ is written as:

$$|x^{(i)}\rangle \langle x^{(i)}| = \sum_{k=1}^N \sum_{m=1}^N x_k^{(i)} x_m^{(i)} |k\rangle \langle m| \quad (160)$$

which in a matrix notation is represented by:

$$|x^{(i)}\rangle \langle x^{(i)}| = \begin{bmatrix} x_1^{(i)} x_1^{(i)} & x_1^{(i)} x_2^{(i)} & \cdots & x_1^{(i)} x_N^{(i)} \\ x_2^{(i)} x_1^{(i)} & x_2^{(i)} x_2^{(i)} & \cdots & x_2^{(i)} x_N^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ x_N^{(i)} x_1^{(i)} & x_N^{(i)} x_2^{(i)} & \cdots & x_N^{(i)} x_N^{(i)} \end{bmatrix}. \quad (161)$$

Thus, the sum over training examples produces the following matrix:

$$\frac{1}{M} \sum_{i=1}^M |x^{(i)}\rangle \langle x^{(i)}| = \frac{1}{M} \begin{bmatrix} \sum_i x_1^{(i)} x_1^{(i)} & \sum_i x_1^{(i)} x_2^{(i)} & \cdots & \sum_i x_1^{(i)} x_N^{(i)} \\ \sum_i x_2^{(i)} x_1^{(i)} & \sum_i x_2^{(i)} x_2^{(i)} & \cdots & \sum_i x_2^{(i)} x_N^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i x_N^{(i)} x_1^{(i)} & \sum_i x_N^{(i)} x_2^{(i)} & \cdots & \sum_i x_N^{(i)} x_N^{(i)} \end{bmatrix} \quad (162)$$

which for demeaned data is equivalent to covariance matrix. The density matrix corresponding to correlation matrix could be obtained if the classical data is also standardized in the first step of algorithm.

Exponential of density matrix

Quantum phase estimation subroutine *PhaseEstim* can be used to obtain eigenvectors and eigenvalues of the corresponding density matrix ρ . However, the quantum subroutine *PhaseEstim* requires unitary matrix U as an input, whereas generally density matrix ρ does not meet that condition. In order to solve that problem we use mathematical law stating that $U = e^{iH}$ is unitary for any Hermitian matrix H . Obviously, density matrix ρ is Hermitian by definition, so the only thing that remains is to take that density matrix and build the unitary gate $U = e^{i\rho}$. The knowledge of how physically exponentiate density matrix is not required to understand the algorithm, however the bulk of the algorithm consists of the ability to efficiently generate the exponent of an arbitrary density matrix ρ . The method is presented in Ref. [28] and for the purpose of this algorithm it should be mentioned that the actual exponent generated $e^{-i\rho t}$ is with additional factor t and the time complexity of generating this matrix is $\mathcal{O}(\log N)$.

Eigendecomposition of density matrix

Having generated unitary matrix $U = e^{-i\rho t}$ we can apply quantum phase estimation subroutine *PhaseEstim*. The additional factor t in the exponent of unitary matrix would be helpful as in quantum phase estimation algorithm we use powers of unitary gate U . The eigenvectors of ρ are also eigenvectors of $e^{-i\rho t}$ and the eigenvalues λ of ρ are just exponentiated $e^{-i\lambda t}$. The standard quantum phase estimation takes unitary gate U as well as one of its eigenvector $|\phi\rangle$ and does the transformation:

$$|0\rangle^{\otimes n} |\phi\rangle \rightarrow |2^n \theta\rangle |\phi\rangle \quad (163)$$

with phase θ in a control qubit allowing to calculate eigenvalue $\lambda = e^{2\pi i \theta}$ and n denoting number of control qubits storing estimate of phase θ . The qPCA algorithm uses slightly different phase estimation algorithm by applying unitary matrix $U = e^{-i\rho t}$ to density matrix ρ instead of eigenvector. To see how it works, we firstly present the result of unitary matrix application on one of the pure state $|x^{(i)}\rangle$ from mixed state described by density matrix ρ :

$$\begin{aligned} e^{-i\rho t} |x^{(i)}\rangle &= \sum_{j=1}^M e^{-i\lambda^{(j)} t} |\phi^{(j)}\rangle \langle \phi^{(j)} | x^{(i)} \rangle \\ &= \sum_{j=1}^M e^{-i\lambda^{(j)} t} \langle \phi^{(j)} | x^{(i)} \rangle |\phi^{(j)}\rangle \\ &= \sum_{j=1}^M c^{(ij)} |\phi^{(j)}\rangle \end{aligned} \quad (164)$$

with $c^{(ij)} = e^{-i\lambda^{(j)} t} \langle \phi^{(j)} | x^{(i)} \rangle$. Then the output of quantum phase estimation algorithm is a superposition:

$$|0\rangle^{\otimes n} |x^{(i)}\rangle \rightarrow \sum_{j=1}^M c^{(ij)} |\tilde{\lambda}^{(j)}\rangle |\phi^{(j)}\rangle. \quad (165)$$

Due to the fact that we use unitary matrix $e^{-i\rho t}$ with eigenvalues $e^{-i\lambda^{(j)} t}$, the control state after *PhaseEstim* subroutine equals $\tilde{\lambda}^{(i)} = 2^n \frac{\lambda^{(i)} t}{2\pi}$ from which we can directly calculate eigenvalue $\lambda^{(i)}$. The output state can be also written in a density matrix representation as:

$$\eta^{(i)} = \sum_{j=1}^M |c^{(ij)}|^2 |\tilde{\lambda}^{(j)}\rangle \langle \tilde{\lambda}^{(j)}| \otimes |\phi^{(j)}\rangle \langle \phi^{(j)}|. \quad (166)$$

In qPCA algorithm we apply the same logic, but instead of the pure state $|x^{(i)}\rangle$ we use mixed state ρ described by density matrix $\frac{1}{M} \sum_{i=1}^M |x^{(i)}\rangle \langle x^{(i)}|$. Weighting possible outputs $\eta^{(i)}$ of phase estimation algorithm with mixed state probabilities $\frac{1}{M}$ yields the final density matrix:

$$\begin{aligned} \eta &= \sum_{j=1}^M \sum_{i=1}^M \frac{1}{M} |c^{(ij)}|^2 |\tilde{\lambda}^{(j)}\rangle \langle \tilde{\lambda}^{(j)}| \otimes |\phi^{(j)}\rangle \langle \phi^{(j)}| \\ &= \sum_{j=1}^M \lambda^{(j)} |\tilde{\lambda}^{(j)}\rangle \langle \tilde{\lambda}^{(j)}| \otimes |\phi^{(j)}\rangle \langle \phi^{(j)}|. \end{aligned} \quad (167)$$

The $\lambda^{(j)}$ coefficient is derived as follows:

$$\begin{aligned}
\sum_{i=1}^M \frac{1}{M} |c^{(ij)}|^2 &= \sum_{i=1}^M \frac{1}{M} e^{-i\lambda^{(j)}t} \langle \phi^{(j)} | x^{(i)} \rangle \langle x^{(i)} | \phi^{(j)} \rangle e^{i\lambda^{(j)}t} \\
&= \langle \phi^{(j)} | \left(\sum_{i=1}^M \frac{1}{M} |x^{(i)}\rangle \langle x^{(i)}| \right) | \phi^{(j)} \rangle \\
&= \langle \phi^{(j)} | \rho | \phi^{(j)} \rangle \\
&= \lambda^{(j)}.
\end{aligned} \tag{168}$$

To summarize, the difference is that in standard phase estimation algorithm an input is known eigenvector $|\phi\rangle$, whereas in *PhaseEstim* subroutine used by qPCA an input ρ is a mixed state with M unknown eigenvectors $|\phi^{(i)}\rangle$. The time complexity of quantum phase estimation algorithm is comparable to exponentiating of density matrix thus does not increase the time complexity of algorithm. A quantum speed-up comes from the representation of classical data as qubits.

Sampling

Sampling from the final state derived in Eq. (167) allows to reveal features of eigenvectors. Assume, that R principal components sufficiently describe the variance of the data or equivalently the sum of eigenvalues corresponding to these principal components is close to 100%. The sampling of final state yields eigenvector $|\phi^{(j)}\rangle$ and corresponding eigenvalue $\lambda^{(j)}$ with probability $\lambda^{(j)}$. In other words, preparing m copies of final state we would anticipate that the eigenvector $\phi^{(j)}$ will be sampled on average $m\lambda^{(j)}$ times. In case $m \ll R$ and providing that building the exponent of ρ has $\mathcal{O}(\log N)$ complexity, the time complexity of sampling is $\mathcal{O}(R \log N)$. The sampled principal components in quantum states are then used to calculate scores or to reveal other features.

Principal Components and Scores

The following method is proposed to reveal the scores. The scores could be extracted from sampled principal component $|\phi^{(j)}\rangle$ by calculating the projection of an eigenvector on a training vector:

$$\begin{aligned}
\langle x^{(i)} | \phi^{(j)} \rangle &= \sum_{k=1}^N \sum_{m=1}^N x_k^{(i)} \phi_m^{(j)} \langle k | m \rangle \\
&= \sum_{k=1}^N x_k^{(i)} \phi_k^{(j)} \\
&= s_i^{(j)}
\end{aligned} \tag{169}$$

for each $i = \{1, \dots, M\}$ training vector encoded in quantum state and thus recreating score for j -th principal component:

$$S^{(j)} = [s_1^{(j)}, s_2^{(j)}, \dots, s_M^{(j)}]^T \tag{170}$$

The overlap $\langle x^{(i)} | \phi^{(j)} \rangle$ could be calculated by the quantum subroutine *SwapTest*. The R first scores corresponding to the highest eigenvalues represent the compressed data and can be used in another machine learning algorithm. Additionally, we can reveal additional features of sampled eigenvector $|\phi^{(j)}\rangle$ by measuring the expectation value $\langle \phi^{(j)} | M | \phi^{(j)} \rangle$ for an arbitrary observable M . The other machine learning algorithms could be accelerated if they can use these features in their set-up. Note that the qPCA algorithm is efficient in case $R \ll N$, meaning that a relatively small number of principal components are required to explain variance of data. In other cases the time complexity of algorithm is similar to the classical one.

3. Summary

The qPCA algorithm is based on the eigendecomposition of a density matrix ρ and it is executed by *PhaseEstim* subroutine. The sampling of a final state provides us principal components in quantum states from which we can calculate scores or reveal other features of eigenvectors. The time complexity of the algorithm consists of exponentiating density matrix ρ in $\mathcal{O}(\log N)$ and sampling a final state which results in the overall time complexity $\mathcal{O}(R \log N)$. This

is an exponential speedup comparing to $\mathcal{O}(N)$ time complexity for classical PCA. The qPCA speed-up applies only if a data could be explained by a relatively small number of principal components. The quantum speed-up is caused mainly by the representation of classical data as qubits.

E. Quantum Neural Networks

1. Preliminaries

The classical neural networks algorithm uses the layers of connected neurons and selects the weights in each layer so that the cost function is minimized. The neurons are represented by activation functions such as sigmoid or softmax functions, which introduce the nonlinearity to the algorithm. The neural networks algorithm is mostly used in pattern recognition, classification problems and financial time series prediction in which they try to learn the non-linear dependencies between an input and output data. Currently, there is no specific quantum version of the algorithm, much work is concentrated on Hopfield neural networks, which are more close to neuroscience than to machine learning. In this section, the quantum feedforward neural networks (qNN) is presented as it is easier in the explanation. Moreover, the classical feedforward neural networks are very popular in machine learning applications. The qNN algorithm is based on the work of Wan et al. [29] and is a straightforward generalization of classical feedforward neural network. The logic behind the algorithm is to turn classical components of a neural network to quantum components in a step by step manner.

2. Algorithm

Classical feedforward neural network

We start with classical feedforward neural network. For simplicity, we assume that the input vector is only two dimensional $N = 2$, characterized by classification label y and there is only one hidden layer with one neuron. This could be represented by the graph:

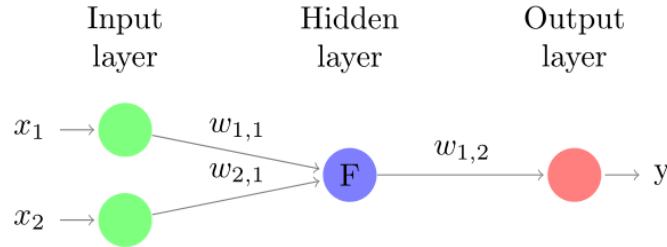


FIG. 18: Classical feedforward neural network

with $w_{i,l}$ indicating i -th input to layer l weight and the activation function denoted by F . Immediately, two problems arise with regard to quantum computations:

- The quantum operations need to be reversible and unitary. The classical activation function F is not reversible as it squeezes the two dimensional vector to a single number y ,
- The quantum states multiplied by weights would not have any practical implications as states are always normalized.

Reversible feedforward neural network

In order to fix irreversibility of classical neural networks, an extra variable denoted by 0 is added to the input vector and the activation function F is changed to output x_1 , x_2 and y . To summarize, the function F has three inputs and three outputs making the operation reversible. The graph presents the approach:

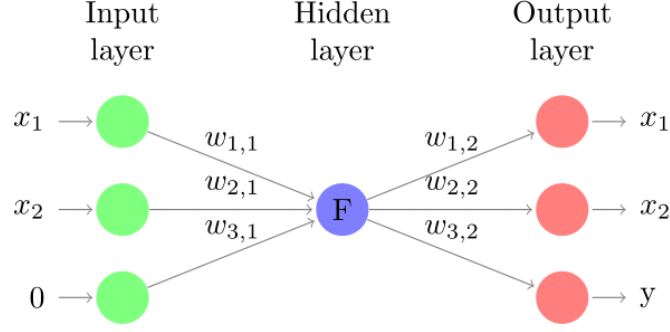


FIG. 19: Reversible classical feedforward neural network

Representation of classical data in quantum states

We would like to represent classical 2-dimensional feature vector $x = [x_1, x_2]^T$ together with label y as quantum states. It can be achieved by representing classical data as binary strings:

$$\begin{aligned} x_1 &\rightarrow [a_{1,1} \ a_{1,2} \ \dots \ a_{1,k}]^T \\ x_2 &\rightarrow [a_{2,1} \ a_{2,2} \ \dots \ a_{2,k}]^T \\ y &\rightarrow [b_1 \ b_2 \ \dots \ b_m]^T \end{aligned} \quad (171)$$

with $a_{i,j} = \{0, 1\}$ for $i = 1, 2$ and $j = 1 \dots, k$ as well as $b_p = \{0, 1\}$ for $p = 1 \dots, m$. We assumed that each entry of feature vector could be expressed in k bits and label could be represented by m bits. The binary strings can be then directly translated into k qubit states for feature vector entries and m qubit state for label:

$$\begin{aligned} |x_1\rangle &= |a_{1,1}\rangle \otimes |a_{1,2}\rangle \otimes \dots \otimes |a_{1,k}\rangle \\ |x_2\rangle &= |a_{2,1}\rangle \otimes |a_{2,2}\rangle \otimes \dots \otimes |a_{2,k}\rangle \\ |y\rangle &= |b_1\rangle \otimes |b_2\rangle \otimes \dots \otimes |b_m\rangle \end{aligned} \quad (172)$$

Different ways could be chosen to represent a classical data in quantum states and the efficiency of the algorithm could be different for each method applied.

Quantum feedforward neural network

The last step is to change reversible function F into an unitary quantum gate U , which inputs and outputs are quantum states. The graph below illustrates the final approach:

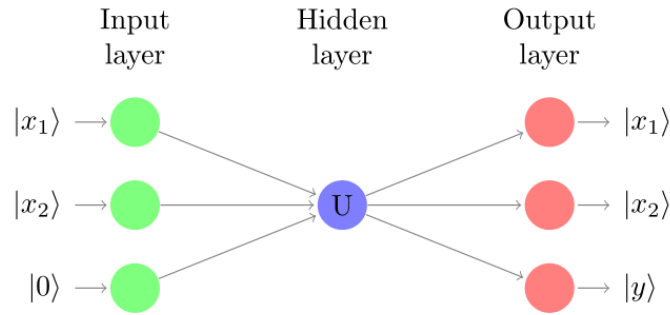


FIG. 20: Quantum feedforward neural network

An arbitrary unitary matrix acting on n qubits can be represented as:

$$U = \exp \left[i \left(\sum_{k_1, \dots, k_n=0, \dots, 0}^{3, \dots, 3} \alpha_{k_1, \dots, k_n} (\sigma_{k_1} \otimes \dots \otimes \sigma_{k_n}) \right) \right] \quad (173)$$

where σ_k are Pauli matrices for $k = \{1, 2, 3\}$ and σ_0 is just 2x2 identity matrix. The parameters α_{k_1, \dots, k_n} are trainable parameters which can be obtained minimizing the cost function. This is different approach in comparison to classical neural networks as the activation functions (instead of numeric weights) represented by unitary

matrices are trainable. The representation can be easily generalized to unitary matrix acting on any number of qubits.

Minimization of cost function

The cost function C is proposed to be a sum of overlays over M training examples (multiplied by -1 to reach solution by minimization). This is written as:

$$C = - \sum_{j=1}^M \langle y_{model}^{(j)} | y^{(j)} \rangle. \quad (174)$$

The modeled $y_{model}^{(j)}$ is dependent on unitary matrix parameters α_{k_1, \dots, k_n} , thus in each optimization iteration the problem reduces to selection of proper values α_{k_1, \dots, k_n} , build a new quantum gate U and calculate the value of cost function by retrieving the overlay $\langle y_{model}^{(j)} | y^{(j)} \rangle$ for each training example j . The gradient descent can be chosen as minimization procedure, updating the parameters α_{k_1, \dots, k_n} as follows:

$$\delta \alpha_{k_1, \dots, k_n} = -\eta \frac{\partial C}{\partial \alpha_{k_1, \dots, k_n}} \quad (175)$$

with learning parameter η controlling how much the coefficients are changing in each update. Obviously, the discretized version of gradient descent should be used. Alternatively, the paper [29] defines the cost function as:

$$C = \sum_{j=1}^M \sum_{i=1}^3 f_{i,j} (\langle \sigma_i \rangle_{model} - \langle \sigma_i \rangle)^2 \quad (176)$$

where $f_{i,j}$ are non-negative real parameters and $\langle \sigma_i \rangle$ are the expectation values of Pauli matrices on individual outputs $y^{(j)}$. The Pauli matrices $\sigma_1, \sigma_2, \sigma_3$ are observables of particle spin in x, y and z directions, respectively. Thus, the cost function expresses the difference between modeled labels and actual labels by measuring the observables on states $y_{model}^{(j)}$ and $y^{(j)}$ and comparing the average results. The more modeled labels and actual labels differs, the bigger will be the difference between expectation values, thus it is a valid definition of cost function.

3. Summary

The benefits of quantum neural networks in comparison to the classical algorithm are not yet specified. A potential speed-up can come from a superposition, however the assessment of unitary matrix learning speed is difficult. The quantum algorithm presented is, in fact, classical in terms of learning parameters, so possible further improvement can be achieved by representing α_{k_1, \dots, k_n} parameters in a quantum state allowing for superposition. There are also concerns how the unitary matrix learning will be implemented experimentally. Nevertheless, the quantum neural networks algorithm is an interesting and promising field of quantum machine learning algorithms that is currently under development.

VI. CONCLUSIONS

The work presented the key elements required to understand the quantum machine learning algorithms. The basic quantum theory, quantum computation and quantum algorithms formed a proper basis for the introduction of quantum machine learning algorithms. The author of this paper tries to simplify and efficiently pass the knowledge by grouping quantum algorithms into quantum subroutines and explaining them in a step-by-step manner with examples. The paper focuses on selected supervised and unsupervised machine learning algorithms and carefully describes their quantum versions. Nevertheless, the paper is by no means the complete review, further quantum algorithms such as quantum decision tree [30], quantum Bayesian methods [31, 32] and quantum associative memory [33] exist in the literature. Moreover, the quantum annealing techniques implemented by adiabatic quantum computing are developed [22, 34] and can be used as an alternative to quantum circuit approach. Most of the algorithms presented in the paper provided a significant speed-up in comparison to their classical counterparts, making it a promising field of quantum computation that can enhance the area of machine learning. On the other hand, one should remember that quantum circuit algorithms presented in the paper are sensitive to a decoherence, thus any noise present in a quantum system may destroy the benefits of quantum computing. Although the assessment of time complexities is conducted, the exact execution time is not certainly known until the algorithms would be run on a quantum computer. To sum up, even

though there is much work required in theory and implementation, the quantum machine learning algorithms may solve problems that today's data scientists and company's management are facing in terms of time and calculation resources. Hopefully, the paper is just another step to accomplish this goal by making the knowledge about quantum algorithms more accessible.

-
- [1] M. Hilbert and P. López, The world's technological capacity to store, communicate, and compute information, *Science*, 332(6025) (2011) pp. 60–65.
 - [2] IBM Builds Its Most Powerful Universal Quantum Computing Processors, <https://www-03.ibm.com/press/us/en/pressrelease/52403.wss> (2017).
 - [3] L. de Broglie, XXXV. A tentative theory of light quanta, *Phil. Mag* 47(278) (1924) pp. 446–458.
 - [4] C. J. Davisson, L. H. Germer, Reflection of Electrons by a Crystal of Nickel, *Proceedings of the National Academy of Sciences of the United States of America* 14(4) (1928) pp. 317–322.
 - [5] A. Einstein, Concerning an Heuristic Point of View Toward the Emission and Transformation of Light, *Ann. Phys.* 17 (1905) 132.
 - [6] M. Born, *The Born Einstein Letters* (Walker and Company, New York, 1971).
 - [7] D. J. Griffiths, *Introduction to Quantum Mechanics* (Cambridge University Press, 2016).
 - [8] Quirk, <http://algassert.com/quirk>.
 - [9] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP 2: A Python framework for the dynamics of open quantum systems, *Comp. Phys. Comm.* 184 (2013) 1234.
 - [10] L. K. Grover, A fast quantum mechanical algorithm for database search, *arXiv preprint quant-ph/9605043* (1996).
 - [11] C. H. Bennett, E. Bernstein, G. Brassard, U. Vazirani, The strengths and weaknesses of quantum computation, *SIAM Journal on Computing*, 26(5) (1997) pp. 1510–1523.
 - [12] P. J. Coles et al. Quantum Algorithm Implementations for Beginners, *arXiv:1804.03719* (2018).
 - [13] C. Dürr and P. Hoyer, A quantum algorithm for finding the minimum, *arXiv preprint quant-ph/9607014* (1996).
 - [14] A. Ahuja and S. Kapoor, A Quantum Algorithm for finding the Maximum, *arXiv preprint quant-ph/9911082* (1999).
 - [15] W. P. Baritmpa, D. W. Bulger and G. R. Wood, Grover's Quantum Algorithm Applied to Global Optimization, *SIAM Journal on Optimization*, 15(4) (2005) pp. 1170–1184
 - [16] E. Farhi, J. Goldstone and S. Gutmann, A Quantum Approximate Optimization Algorithm, *arXiv preprint quant-ph/1411.4028* (2014).
 - [17] G. Verdon, M. Broughton and J. Biamonte, A quantum algorithm to train neural networks using low-depth circuits, *arXiv preprint quant-ph/1712.05304* (2017).
 - [18] L. Hales and S. Hallgren, An improved quantum Fourier transform algorithm and applications, *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, (2000) pp. 515.
 - [19] R. Cleve, A. Ekert, C. Macchiavello and M. Mosca, Quantum algorithms revisited, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, (1998) pp. 454.
 - [20] C. D. Manning, P. Raghavan and H. Schtze, *Introduction to Information Retrieval* (Cambridge University Press, New York, 2008).
 - [21] E. Aïmeur, G. Brassard, and S. Gambs, Machine learning in a quantum world, *Advances in Artificial Intelligence*, Springer (2006) pp. 431–442.
 - [22] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum algorithms for supervised and unsupervised machine learning, *arXiv preprint 1307.0411* (2013).
 - [23] P. Niemann, R. Datta, and R. Wille, Logic Synthesis for Quantum State Generation, *IEEE 46th International Symposium on Multiple-Valued Logic*, Springer (2016) pp. 247–252.
 - [24] M. Mttönen, J. Vartiainen, V. Bergholm, and M. M. Salomaa, Transformation of quantum states using uniformly controlled rotations, *Quantum Information and Computation*, 5 (2005) pp. 467–473.
 - [25] E. Aïmeur, G. Brassard, and S. Gambs, Quantum clustering algorithms, *Proc. 24th international conference on machine learning*, (2007) pp. 1–8.
 - [26] D. Anguita, S. Ridella, F. Riviaccio and R. Zunino, Quantum optimization for training support vector machines, *Journal Neural Networks - 2003 Special issue: Advances in neural networks research*, 16(5–6) (2003) pp. 763–770.
 - [27] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum support vector machine for big data classification, *arXiv preprint 1307.0471* (2013).
 - [28] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, *arXiv preprint 1307.0401* (2013).
 - [29] K. H. Wan, O. Dahlsten, H. Kristjánsson, R. Gardner and M.S. Kim, Quantum generalisation of feedforward neural networks, *arXiv preprint 1612.01045* (2016).
 - [30] S. Lu and S. Braunstein, Quantum decision tree classifier, *Quantum Information Processing*, 13(3) (2014) pp. 757–770.
 - [31] M. Guță and W. Kotłowski, Quantum learning: asymptotically optimal classification of qubit states, *New Journal of Physics*, 12(12) (2010) 123032.
 - [32] M. Sasaki, A. Carlini, and R. Jozsa, Quantum template matching, *Physical Review A*, 64(2) (2001) 022317.
 - [33] C. Trugenberger, Quantum pattern recognition, *Quantum Information Processing*, 1(6) (2002) pp. 471–493.
 - [34] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining* (Academic Press, 2014).