



UNIVERSITAT DE
BARCELONA

Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

Efficient unitary approximations
in quantum computing: the
Solovay-Kitaev Theorem

Autor: Josep Lumbreras Zarapico

Director: Dra. Joana Cirici
Realitzat a: Departament
de Matemàtiques
i Informàtica

Barcelona, 18 de gener de 2018

Abstract

Over the past few years, quantum computing has become more plausible due to the great advances in technology. While quantum computers are on their birth, the underlying mathematics have evolved to the point of proving that some quantum algorithms can solve problems that were unsolvable in classic computers. In order to implement these algorithms in a real machine, it is important to develop efficient ways to do it. The Solovay-Kitaev Theorem states that is possible. This work pretends to offer a complete review of the Solovay-Kitaev Theorem giving all the necessary tools to prove it. Moreover, we offer a brief introduction to the standard mathematical model of quantum computing, based on unitary operations.

Resum

En els últims anys, la computació quàntica ha esdevingut més factible a causa dels grans avenços tecnològics. Mentre els ordinadors quàntics encara estan en el seu naixement, les matemàtiques han avançat fins el punt de demostrar que certs algorismes quàntics poden resoldre problemes que eren impossible de resoldre en un ordinador clàssic. Per implementar aquests algorismes en una màquina real, és important desenvolupar formes eficients de fer-ho, i aquí és on el teorema de Solovay-Kitaev prova que és possible. Aquest treball pretén oferir una revisió completa del Teorema de Solovay-Kitaev donant totes les eines necessàries per demostrar-lo. A més, oferim una breu introducció al model matemàtic estàndard de la computació quàntica basat en operacions unitàries.

Acknowledgements

I would like to thank my advisor Joana Cirici, whose advice and corrections have been essential for the development of this text.

I would also like to thank my parents and Julia for their support during all these years.

Contents

1	Introduction	4
2	Preliminaries on Lie groups and Lie algebras	8
2.1	Matrix Lie groups	8
2.2	The Matrix Exponential	11
2.3	Lie Algebras	13
2.4	Matrix Norm	17
2.5	Unitary transformations	18
2.6	The Lie algebra of $SU(2)$	19
2.7	Connecting $SU(2)$ with $SO(3)$	20
3	Quantum computing	22
3.1	Qubit	22
3.2	The density matrix	24
3.3	Bloch Sphere	24
3.4	Single qubit operations	25
3.5	Universal quantum gates	28
4	Solovay-Kitaev	32
4.1	Preliminaries on distances in $SU(2)$	32
4.2	Shrinking Lemma	35
4.3	The Solovay Kitaev Theorem	37
4.4	The initial epsilon net	39
4.5	Generalization to $SU(d)$	40
4.6	Controlling the nets	42
5	Conclusions	44

1 Introduction

“A classical computation is like a solo voice, one line of pure tones succeeding each other. A quantum computation is like a symphony, many lines of tones interfering with one another.”

—Seth Lloyd, *Programming the Universe*.

Nowadays we are living what many scientists call, a second quantum revolution. The first quantum revolution dates back to the first half of the 20th century, when scientists understood the basic rules of quantum mechanics which were the foundation that allowed inventions like the laser or the transistors, the basic building blocks of computers. Over the past few years, technology has developed to the point where we have control over one single atom, which means that quantum properties like superposition or entangling can be used to build new devices, in particular the quantum computer. The first ideas of quantum computing were established in the early eighties, but during the last years the great advances in mathematics, materials science, and computer science have turned quantum computing from a theory into a reality.

The main idea of quantum computing relies on the physical device where the information is stored. Quantum computing uses physical systems like atoms, superconducting circuits or photons, that allow to create superpositions of classical states. For example, an electron can be in two levels of energy, the ground state and an excited state, and in each state we can store information as 0 or 1 (like the bits in classical computing). However, quantum mechanics allow physical states to be in a superposition, so we can have states that are at the same time 0 and 1. More precisely, if we imagine a sphere and we associate the 0 to the north pole and the 1 to the south pole, the quantum state can be any point on the surface of the sphere, and these points is what we call qubits; the analogues to the bit in quantum computing. This way of seeing the qubit as a point on a sphere is more accurate rather than saying that is in both states 0 and 1 at the same time.

The great advantage of quantum computing over classical, is the exponential scaling of the quantum systems in front of the classical. Since a qubit can represent two bits states, n qubits can represent 2^n bits states, and this fact allows to manipulate more information with less resources. There are different physical implementations of a quantum computer (see Chapter 7 of [NC00]), but the key point is that the operations that can be done in a quantum system are unitary transformations. Mathematically, it turns out that quantum computing can be described by vectors in \mathbb{C}^{2^n} representing the qubits, and elements of the unitary group $U(n)$ representing the operations (like the classical gates NOT, XOR. . .). From this new way of computing a new type of algorithms completely different from the classical arise. Also, with these new algorithms, quantum computers may be able to efficiently solve some problems which classical computers could not solve. One of the most promising quantum algorithms is Shor’s algorithm [Mon16], which allows to solve efficiently integer factorization, a problem that classically belongs to the complexity class NP. Other useful applications will

be in fields like machine learning, economical models, artificial intelligence, or molecular simulations, which could help to develop new types of drugs and materials. However, near term quantum computers are restricted by the number of qubits that can handle and the amount of operations that we can perform. Algorithms like Shor's requires thousands of qubits, while the best quantum computers that we have can run less than 50 qubits, so if one wishes to run useful quantum algorithms it is important to find efficient ways to do it.

Now we are approaching what John Preskill called the NISQ era (Noisy Intermediate-Scale Quantum) [Pre18]. In this near stage we will deal with quantum computers that operate between 50 and 100 qubits, not enough for running Shor's algorithm, but in fields like the simulation of many body quantum systems, quantum computers will offer a great advantage over the classical. In this scenario, apart from the limitation on the number of qubits we will have to deal with "noise", which means that the devices will offer errors due to the fragility of the quantum systems. This fact raises two problems in the implementation of quantum algorithms: the first and most obvious is that we can not build quantum circuits with an arbitrary number of quantum operations, since if the computation takes too long, the information can be lost due to the fragility of the system. The second is that we will have to implement error correction to the algorithms in order to avoid problems like the loss of the information of some qubits.

Most quantum algorithms can be decomposed in quantum gates that operate on a single qubit, together with some other gates that act on two qubits (called entangling gates). However, since one will be restricted by the number of gates that can use to operate in a quantum circuit, it is necessary to work with a set of gates that can approximate any unitary operation and can perform fault tolerant computations. A set that can approximate with an arbitrary precision any unitary operation is called a universal set of quantum gates. Finding a universal set of quantum gates is equivalent to finding dense sets in the unitary group $U(n)$ (these facts will be explained at the end of Section 3). Note that since the unitary group $U(n)$ is infinite, it is impossible to achieve any element of the group with a finite set, since one is uncountable and the other countable. This is why a universal set of quantum gates cannot achieve an arbitrary element of $U(n)$ with zero error. With a universal set of quantum gates it is important to know how fast one can achieve any unitary operation with a certain precision, in the sense that we want to construct the operations with short sequences of elements of our gate set. The result that states how fast a universal set of quantum gates can achieve any unitary operation is the Solovay-Kitaev Theorem. In the simplest form, it states that:

Theorem (Solovay-Kitaev). *Given a set of elements in $SU(2)$ that generates a dense subset, then it is possible to find approximations for any element of $SU(2)$ with short sequences of elements of the given set.*

Specifically, the theorem states that the length of the approximation scales polylogarithmically in function of the error of the approximated element. Moreover, this result is important for defining quantum complexity classes. Like in

the classical case, we can define new complexity classes for quantum algorithms, the most known are BQP (bounded-error quantum polynomial time) and QMA (Quantum Merlin Arthur) which are the bounded-error quantum analogues of P and NP. In order to have a good definition for the classes BQP and QMA, we need independence of the gate set that is used for the computation. Here the Solovay-Kitaev Theorem states that a change of a gate set only increases the complexity of the algorithms in a polylogarithmic factor, so they can be well defined. However, there are certain details of the exact statement of the theorem that have to be improved (see [Kup15]).

The aim of this work is to introduce the Solovay Kitaev theorem for readers without prior knowledge in quantum mechanics or quantum computation. The proof of this theorem is based on exploiting the non-abelian structure of $SU(2)$. Basically, it consists on picking a dense subset of $SU(2)$ and generating a good approximation for the identity. Then, using some properties of the distance between the elements of the group, it can be proved that we can achieve better approximations for the identity at the cost of having a more dense subset. For these approximations we will use the group commutator. This is an operator for the unitary group which is equal to the identity whenever two elements commute. Iterating the process of having better approximations for the identity we will have elements very close to the identity, and using a translation step we can move these dense subsets to other points of the group. In this manner, we will achieve approximations for any element of $SU(2)$.

Solovay announced the theorem in 1995 for the case $SU(2)$ in an email discussion but he did not publish it. In 1997, Kitaev generalized the result to the case of $SU(d)$ and published it in a review paper [Kit97]. The proof that we will give is based on the Appendix 3 of [NC00], and it involves some geometrical explanations matching distances between $SU(2)$ and \mathbb{R}^3 . Note that in the present text, we provide all the computations that were left as exercises in the book, some of them taken from [Har01] and [Ozo09]. Also, throughout our exposition we have tried to add all the mathematical rigour that is sometimes lacking in the standard expositions of the Solovay-Kitaev Theorem, most of them intended for physicists and quantum computer scientists.

As we will see, a main ingredient in the proof of the Solovay-Kitaev Theorem is the notion of dense subset in a metric space with a biinvariant metric. With this idea in mind, the theorem has been extended, generalized and adapted to other settings that we do not cover in the present work, such as [AB17],[ND05],[HBC02],[Ozo09]. Also, several prominent topologists and geometers have recently shown their interest and developed works related to the ideas of the original Solovay-Kitaev Theorem (see for instance the work of Freedman-Kitaev-Lurie [FKL03] and Sarnak's lectures on golden gates [Sar15]). Some of the ideas underlying these more general proofs are briefly explained in the last section.

This work is organized into three sections as follows:

Section 2 focuses on the unitary group $U(n)$. In particular, we give some fundamental properties for the case $n = 2$, which will help us understand the

concept of single qubit gates. Also, we provide a brief introduction to matrix Lie groups and their Lie algebras.

In Section 3, we introduce the notion of qubit as a point in a complex projective space. Then, we recall the Bloch sphere representation, which is a geometrical way for understanding the qubit. Using some results of Section 2, we give a relation between a quantum gate and a rotation over the Bloch sphere. Lastly, we define a universal set of quantum gates, a key ingredient of the Solovay-Kitaev Theorem. In this section, we also include a standard proof of a particular set of gates that is universal.

Section 4 is the core of this work: we give a proof of the Solovay Kitaev Theorem for the case of $SU(2)$. Also, we provide all the computations needed to prove the Shrinking Lemma, is an essential result for proving the Solovay Kitaev Theorem. Finally we sketch the proof for the general case $SU(d)$ and give some details of an alternative proof of the theorem.

In Section 5, we collect some final considerations, comment on alternative and improved proofs, and generalizations of the Solovay-Kitaev Theorem.

For Section 2, we will assume that the reader has familiarity with the main notions and properties on differentiable manifolds. Some basic references are [Lan99] and [War83], although we will not need much more than the definitions of differentiable manifold and differentiable map, always within the point of view of Lie groups. For Sections 3 and 4 we do not assume any prerequisites, since we will provide a brief introduction to the mathematics used in quantum mechanics.

To conclude we would like to note that this work does not include any quantum algorithm, quantum circuit or explanations about how it is possible to compute with a quantum computer. We encourage the interested reader to take a look at references like [NC00] or [KN02], where the quantum computational model is well defined. These references include the most representative's quantum algorithms.

2 Preliminaries on Lie groups and Lie algebras

Lie groups appear naturally in almost every theory where the notion of symmetry plays a role. Many systems studied in theoretical physics show some form of symmetry. For instance, in particle physics Lie groups and their associated representation theory are useful to model the symmetries of subatomic particles. In this section, we review some main definitions and properties on Lie groups and Lie algebras that we will use throughout the text. We center our attention to matrix Lie groups and in particular, to the groups $U(n)$ and $SU(n)$ of unitary matrices, whose definition and main properties we recall below. We will mostly follow the book [Hal15] of Hall, which focuses on matrix Lie groups. Other references on the subject are [FH91] and [War83].

2.1 Matrix Lie groups

Definition 2.1. A *Lie group* is a smooth manifold G which is also a group and such that the group product:

$$\mu : G \times G \longrightarrow G$$

and the inverse map are smooth. These conditions can be combined into the single requirement that the map $G \times G \longrightarrow G$ given by $(x, y) \mapsto x^{-1}y$ is smooth.

Definition 2.2. Let G and H be Lie groups. A *Lie group homomorphism* from G to H is a group homomorphism $\Phi : G \rightarrow H$ which is also a smooth map. If, in addition, Φ is one-to-one and onto and the inverse map Φ^{-1} is smooth, then Φ is called a *Lie group isomorphism*.

Examples 2.3. We list some very first examples of Lie groups:

- (a) The euclidean space \mathbb{R}^n is a Lie group under vector addition.
- (b) The non-zero complex numbers \mathbb{C}^* form a Lie group under multiplication.
- (c) The unit circle $S^1 \subseteq \mathbb{C}^*$ forms a Lie group with the multiplication induced from \mathbb{C}^* .

Although many of the results presented in this section are valid over an arbitrary field, to simplify our exposition, from now on we will let \mathbb{K} denote either the field \mathbb{R} of real numbers or the field \mathbb{C} of complex numbers.

Let us now fix some notation on matrices. We will denote by $\mathcal{M}_n(\mathbb{K})$ the set of all $n \times n$ matrices with entries in \mathbb{K} .

- (1) Given $A \in \mathcal{M}_n(\mathbb{K})$, we will denote its (i, j) -entry by A_{ij} , so that

$$A = \begin{pmatrix} A_{00} & \cdots & A_{0n} \\ \vdots & \ddots & \vdots \\ A_{n0} & \cdots & A_{nn} \end{pmatrix}.$$

(2) The *trace* of $A \in \mathcal{M}_n(\mathbb{K})$ will be denoted by $\text{trace}(A)$:

$$\text{trace}(A) := \sum_{i=1}^n A_{ii}.$$

(3) The *transpose* of $A \in \mathcal{M}_n(\mathbb{K})$ will be denoted by A^{tr} :

$$(A^{tr})_{ij} := A_{ji}.$$

(4) The *adjoint* of $A \in \mathcal{M}_n(\mathbb{C})$ will be denoted by A^* :

$$(A^*)_{jk} := A_{kj}^*,$$

where A_{kj}^* denotes the complex conjugate of A_{kj} . If $A^* = A$, A is said to be an *Hermitian* matrix.

(5) The *commutator* of two matrices $A, B \in \mathcal{M}_n(\mathbb{K})$ is given by

$$[A, B] := AB - BA$$

and the *anticommutator* by

$$\{A, B\} := AB + BA.$$

Definition 2.4. The *general linear group over \mathbb{K}* , denoted by $\text{GL}(n; \mathbb{K})$, is the group of all $n \times n$ invertible matrices with entries in \mathbb{K} :

$$\text{GL}(n; \mathbb{K}) := \{A \in \mathcal{M}_n(\mathbb{K}); \det(A) \neq 0\}.$$

Matrix Lie groups are, by definition, closed subgroups of the general linear group. To make this definition precise, let us recall the notion of convergence in the space of matrices.

Definition 2.5. Let A_m be a sequence of complex matrices in $\mathcal{M}_n(\mathbb{C})$. We say that A_m *converges* to a matrix A if each entry of A_m converges (as $m \rightarrow \infty$) to the corresponding entry of A .

Definition 2.6. A *matrix Lie group* is a subgroup G of $\text{GL}(n; \mathbb{C})$ with the following properties:

- (1) The identity matrix is in G .
- (2) For all A and B in G , the matrices AB and A^{-1} are also in G .
- (3) If A_m is any sequence of matrices in G , and A_m converges to some matrix A , then either A is in G or A is not invertible.

Remark 2.7. Many works give these definitions for the specific case of $\mathbb{K} = \mathbb{C}$ since any matrix Lie group should be a subgroup of this case.

Remark 2.8. Every matrix Lie group is a Lie group. To see this, one proves that every matrix Lie group is a smooth embedded submanifold of $M_n(\mathbb{C})$. On the other hand it can be proved that every Lie group is not isomorphic to a matrix Lie group. We will not prove these results, but the curious reader can find the proofs in [Hal15]. Chapter 1 and 2 for the proof of that every matrix Lie group is a Lie group, and Chapter 4 for the other.

Proposition 2.9. *The general linear group $\mathrm{GL}(n; \mathbb{K})$ is a matrix Lie group.*

Proof. If A_m is a sequence of matrices in $\mathrm{GL}(n; \mathbb{C})$ and A_m converges to A , then by definition, either A is in $\mathrm{GL}(n; \mathbb{C})$ or A is not invertible. Moreover, $\mathrm{GL}(n; \mathbb{R})$ is a subgroup of $\mathrm{GL}(n; \mathbb{C})$, and if A_m is a sequence of matrices in $\mathrm{GL}(n; \mathbb{R})$ and converges to A , then the entries of A are real. Thus, either A is not invertible or A is in $\mathrm{GL}(n; \mathbb{R})$. \square

We now review some other interesting matrix Lie groups.

Example 2.10. The *special linear group* over \mathbb{K} is the group of $n \times n$ invertible matrices (with entries in \mathbb{K}) having determinant one:

$$\mathrm{SL}(n; \mathbb{K}) := \{A \in \mathcal{M}_n(\mathbb{K}); \det(A) = 1\}.$$

Note that, since the determinant is a continuous function, if A_n is a sequence of matrices with determinant one and A_n converges to A , then A also has determinant one. This proves that $\mathrm{SL}(n; \mathbb{K})$ is a matrix Lie group.

Example 2.11. The *unitary group*, denoted by $\mathrm{U}(n)$, is the group of $n \times n$ complex matrices $A \in \mathcal{M}_n(\mathbb{C})$ such that:

$$\sum_{\ell=1}^k (A^*)_{j\ell} A_{\ell k} = \delta_{jk},$$

where δ_{jk} denotes the Kronecker delta. From this definition it is easy to check that matrices in $\mathrm{U}(n)$ satisfy $A^*A = AA^* = I$. We have:

$$\mathrm{U}(n) = \{A \in \mathcal{M}_n(\mathbb{C}); A^* = A^{-1}\}.$$

Note that if A belongs to $\mathrm{U}(n)$, then $\det(A) = e^{i\varphi}$ with $\varphi \in \mathbb{R}$.

Example 2.12. The *special unitary group* $\mathrm{SU}(n)$ is the subgroup of $\mathrm{U}(n)$ consisting of unitary matrices with determinant one:

$$\mathrm{SU}(n) := \{A \in \mathcal{M}_n(\mathbb{C}); A^* = A^{-1}, \det(A) = 1\}.$$

In fact, the determinant induces a group homomorphism from $\mathrm{U}(n)$ and $\mathrm{U}(1)$,

$$\det : \mathrm{U}(n) \longrightarrow \mathrm{U}(1).$$

The kernel of this homomorphism is precisely $\mathrm{SU}(n)$.

Example 2.13. The *orthogonal group* $O(n)$ is the group of $n \times n$ real matrices $A \in \mathcal{M}_n(\mathbb{R})$ such that:

$$\sum_{\ell=1}^n A_{\ell j} A_{\ell k} = \delta_{jk}$$

This condition is equivalent to $A^{tr} = A^{-1}$ and it holds if and only if A preserves the inner product on \mathbb{R}^n . We have:

$$O(n) := \{A \in GL(n; \mathbb{R}); A^{tr} = A^{-1}\}.$$

Note that if A belongs to $O(n)$ then $\det(A) = \pm 1$.

Example 2.14. The *special orthogonal group* $SO(n)$ is the subgroup of $O(n)$ given by those matrices with determinant one:

$$SO(n) := \{A \in GL(n; \mathbb{R}); A^{tr} = A^{-1}, \det(A) = 1\}.$$

Geometrically the elements of $O(n)$ are rotations or combinations of rotations and reflections, and the elements of $SO(n)$ are rotations. Since both groups are closed subgroups of $GL(n; \mathbb{C})$, it is clear that they form matrix Lie groups.

2.2 The Matrix Exponential

Before introducing the notion of Lie algebra, we briefly recall some results on the exponential of a matrix. As we will see, this operation plays a crucial role in the theory of Lie groups, connecting matrix Lie groups with their corresponding Lie algebras. Extended details and proofs for the results in this section can be found, for instance, in Section 2 of [Hal15].

Definition 2.15. The *exponential* of a matrix $A \in \mathcal{M}_n(\mathbb{K})$ is defined by:

$$e^A = \sum_{m=0}^{\infty} \frac{A^m}{m!}$$

Sometimes we will denote $\exp(A)$ instead of e^A .

Remark 2.16. As usual A^0 is defined to be the identity matrix and A^m is the product of A , m times with itself.

The main objective of this section is to give some properties of the exponential of a matrix. These results will help us throughout this work, especially in Section 2.3 on Lie algebras and in Section 3.4 on single qubit operations.

Proposition 2.17. *Let $A, B \in \mathcal{M}_n(\mathbb{K})$. Then:*

- (1) $e^0 = I$.
- (2) $(e^A)^* = e^{A^*}$.
- (3) e^A is invertible and $(e^A)^{-1} = e^{-A}$.

(4) If $[A, B] = 0$, then $e^{A+B} = e^A e^B = e^B e^A$.

(5) If C is invertible, then $e^{CAC^{-1}} = C e^A C^{-1}$.

Proof. Point 1 is trivial and Point 2 follows from $(A^m)^* = (A^*)^m$ and the series expansion of the exponential. Point 3 is a special case of Point 4. To check Point it suffices to compute the power series term by term

$$e^A e^B = \left(I + A + \frac{A^2}{2!} + \cdots\right) \left(I + B + \frac{B^2}{2!} + \cdots\right).$$

Rearranging all the terms,

$$e^A e^B = \sum_{m=0}^{\infty} \sum_{k=0}^m \frac{A^k}{k!} \frac{B^{m-k}}{(m-k)!} = \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k=0}^m \frac{m!}{k!(m-k)!} A^k B^{m-k}.$$

Using that A and B commute,

$$(A + B)^m = \sum_{k=0}^m \frac{m!}{k!(m-k)!} A^k B^{m-k}$$

Thus,

$$e^A e^B = \sum_{m=0}^{\infty} \frac{1}{m!} (A + B)^m = e^{A+B}.$$

Point 5 follows from $(CAC^{-1})^m = (CA^m C^{-1})$ and the series expansion of the matrix exponential. □

Given $A \in \mathcal{M}_n(\mathbb{K})$, we may view $e^t A$ as a smooth curve in $\mathcal{M}_n(\mathbb{K})$. We have:

Proposition 2.18. *Given $A \in \mathcal{M}_n(\mathbb{K})$, the following identities are satisfied:*

1. $\frac{d}{dt} e^{tA} = A e^{tA} = e^{tA} A$.
2. $\det(e^A) = e^{\text{trace}(A)}$.

Proof. The first statement is a simple computation. For the second one it suffices to see that if A is diagonalizable with eigenvalues $\lambda_1, \dots, \lambda_n$ then e^A is diagonalizable with eigenvalues $e^{\lambda_1}, \dots, e^{\lambda_n}$. Thus

$$\det(e^A) = e^{\lambda_1} \cdots e^{\lambda_n} = e^{\lambda_1 + \cdots + \lambda_n} = e^{\text{trace}(A)}.$$

If A is not diagonalizable we can distinguish if A is nilpotent or A is arbitrary.

If A is nilpotent means that $N^k = 0$ for k sufficiently large. Then A can be written as $A = C^{-1} X C$ for some invertible matrix C and X an upper diagonal matrix with all the diagonal entries equal to 0. Using Point 5 of Proposition 2.17 the result follows.

If A is arbitrary then it may be written as $A = S + N$, with $[S, N] = 0$, where S is a diagonalizable matrix and N is a nilpotent matrix (see Appendix B.3 of [Hal15]). In particular, its exponential is a finite sum and hence has an explicit expression. Since $[S, N] = 0$, by Proposition 2.17 and the previous cases we have

$$e^A = e^{S+N} = e^S e^N.$$

It now suffices to use the expressions for e^S and e^N respectively. \square

We will also use the following result:

Proposition 2.19. *Let $x \in \mathbb{R}$ and $A \in \mathcal{M}_n(\mathbb{K})$ such that $A^2 = -I$. Then:*

$$e^{iAx} = \cos(x)I + i \sin(x)A.$$

Proof. Using the usual series expansion for the sine and cosine we can compute directly:

$$\begin{aligned} e^{iAx} &= \sum_{m=0}^{\infty} \frac{(iAx)^m}{m!} = \sum_{m=0}^{\infty} \frac{1}{(2m)!} (iAx)^{2m} + \sum_{m=0}^{\infty} \frac{1}{(2m+1)!} (iAx)^{2m+1} \\ &= I \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m)!} x^{2m} + iA \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1)!} x^{2m+1} = \cos(x)I + i \sin(x)A. \end{aligned}$$

\square

2.3 Lie Algebras

Lie algebras are an essential tool in the study of Lie groups. Lie algebras are simpler than Lie groups, because a Lie algebra is a linear space (so they can be understood doing linear algebra). There are some correspondences between a matrix Lie group and a Lie algebra, thus many problems that are hard to work in matrix Lie groups become easier if we work in the Lie algebra.

Definition 2.20. A *Lie algebra over \mathbb{K}* is a \mathbb{K} -vector space \mathfrak{g} , together with a binary operation

$$[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \longrightarrow \mathfrak{g}$$

with the following properties:

(1) $[\cdot, \cdot]$ is bilinear:

$$[aX + bY, Z] = a[X, Z] + b[Y, Z]$$

$$[Z, aX + bY] = a[Z, X] + b[Z, Y]$$

for all $a, b \in \mathbb{K}$ and for all $X, Y, Z \in \mathfrak{g}$.

(2) $[\cdot, \cdot]$ is skew symmetric: $[X, Y] = -[Y, X]$ for all $X, Y \in \mathfrak{g}$.

(3) The *Jacobi identity* holds:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$$

for all $X, Y, Z \in \mathfrak{g}$.

Remark 2.21. A Lie algebra \mathfrak{g} is said to be *commutative* (or *abelian*) if for all $X, Y \in \mathfrak{g}$ we have $[X, Y] = 0$. Any vector space is a commutative Lie algebra when endowed with the trivial bracket, but in general, Lie algebras are not commutative. Also, Lie algebras are not associative in general, but the Jacobi identity is in fact a substitute for associativity.

Example 2.22. Let $\mathfrak{g} = \mathbb{R}^3$ and let $[\cdot, \cdot] : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be given by $[x, y] := x \times y$ where $x \times y$ is the vector product. Then \mathfrak{g} is a Lie algebra. Indeed, bilinearity, skew symmetry, and Jacobi are known properties of the vector product.

Example 2.23. Let \mathcal{A} be an associative algebra and let \mathfrak{g} be a subspace of \mathcal{A} such that $XY - YX \in \mathfrak{g}$ for all $X, Y \in \mathfrak{g}$. Then \mathfrak{g} is a Lie algebra with the binary operation (referred as the *Lie bracket or commutator*) given by

$$[X, Y] := XY - YX.$$

This example will be important for us, since we will work with matrix Lie groups, for which the commutator of matrices defines the binary operation in the Lie algebra associated to such groups.

Every Lie group has an associated Lie algebra. In general, this is defined as the tangent space of the Lie group at the identity. However, the Lie algebra associated to a matrix Lie group admits a simpler and explicit definition via the exponential matrix as we next explain. We conveniently choose this approach, since it will be most useful for our purposes.

Definition 2.24. Let G be a matrix Lie group. The *Lie algebra of G* , denoted by \mathfrak{g} , is the set of all matrices A such that e^{tA} is in G for all real numbers t :

$$\mathfrak{g} := \{A \in \mathcal{M}_n(\mathbb{C}); e^{tA} \in G \text{ for all } t \in \mathbb{R}\}.$$

The Lie bracket of \mathfrak{g} is given by the commutator of matrices.

Theorem 2.25. Let G be a matrix Lie group with Lie algebra \mathfrak{g} . If X and Y are elements of \mathfrak{g} , the following results hold:

- (1) $AXA^{-1} \in \mathfrak{g}$ for every $A \in G$.
- (2) $sX \in \mathfrak{g}$ for every real number s .
- (3) $X + Y \in \mathfrak{g}$.
- (4) $XY - YX \in \mathfrak{g}$.

Idea of proof. Point 1 is easily verified using Proposition 2.17, $e^{tAXA^{-1}} = Ae^{tX}A^{-1}$ and e^tX is in G since X is an element of \mathfrak{g} . Point 2 is immediate since $e^{t(sX)} = e^{(ts)X}$. If A and B commute we have $e^{t(X+Y)} = e^{tX}e^{tY}$ and point 3 is verified. If A and B does not commute we will need to use the *Lie product formula*,

$$e^{t(X+Y)} = \lim_{m \rightarrow \infty} (e^{tX/m}e^{tY/m})^m,$$

and the definition of matrix Lie group (see Chapter 2 of [Hal15] for more details). Point 4 is proved using Proposition 2.18 and the product rule,

$$\frac{d}{dt}(e^{tX}Ye^{-tX})|_{t=0} = (XY)e^0 - e^0YX = XY - YX$$

From Point 1 $e^{tX}Ye^{-tX}$ is in \mathfrak{g} and the result follows using the limit definition of the derivate and noting that \mathfrak{g} is a closed subset of $\mathcal{M}_n(\mathbb{C})$. \square

Like in the previous section, we will now see some useful examples of Lie algebras. In fact, we will give the Lie algebras associated to the matrix Lie groups that we previously introduced.

Remark 2.26. Physicists often use the map $t \mapsto e^{itX}$, so the expressions for the Lie algebras differ by a factor of i from the mathematicians expressions.

Example 2.27. The Lie algebra of $\mathrm{GL}(n; \mathbb{K})$, denoted by $\mathfrak{gl}(n, \mathbb{K})$, is the space $\mathcal{M}_n(\mathbb{K})$ of all $n \times n$ matrices with entries in \mathbb{K} :

$$\mathfrak{gl}(n; \mathbb{K}) = \mathcal{M}_n(\mathbb{K}).$$

Indeed, by Proposition 2.17, the exponential e^{tA} is an invertible matrix for any $t \in \mathbb{R}$ and any $A \in \mathcal{M}_n(\mathbb{C})$. This shows the above identity in the complex case. The real case follows similarly, by noting that if e^{tA} is a real matrix for all $t \in \mathbb{R}$, then A is also a real matrix.

Example 2.28. The Lie algebra of $\mathrm{SL}(n; \mathbb{K})$ denoted by $\mathfrak{sl}(n; \mathbb{K})$ consists of all $n \times n$ complex matrices with trace zero:

$$\mathfrak{sl}(n; \mathbb{K}) = \{A \in \mathcal{M}_n(\mathbb{K}); \text{trace}(A) = 0\}$$

Indeed, if $A \in \mathcal{M}_n(\mathbb{C})$ has trace zero, then by Proposition 2.18 we have $\det(e^{tA}) = 1$. Therefore A belongs to the Lie algebra of $\mathrm{SL}(n; \mathbb{C})$. Conversely, if $\det(e^{tA}) = e^{t \cdot \text{trace}(A)} = 1$ for every real number t , then:

$$\text{trace}(A) = \frac{d}{dt}e^{t \cdot \text{trace}(A)}|_{t=0} = 0$$

Following similar arguments we find that the Lie algebra of $\mathrm{SL}(n; \mathbb{R})$ consists of all $n \times n$ matrices with trace zero.

Example 2.29. The Lie algebra of $U(n)$ denoted by $\mathfrak{u}(n)$ consists of all complex matrices satisfying:

$$\mathfrak{u}(n) = \{A \in \mathcal{M}_n(\mathbb{C}); A^* = -A\}$$

Indeed, a matrix U belongs to $U(n)$ if and only if $U^* = U^{-1}$. Thus using 2.17, e^{tX} is in $U(n)$ if and only if:

$$(e^{tA})^* = (e^{tA})^{-1} = e^{-tA}$$

Again using 2.17 $(e^{tA})^* = e^{tA^*}$ and:

$$e^{tA^*} = e^{-tA}$$

This condition holds for all real t if and only if $A^* = -A$. Thus, this was the condition that we were claiming for A belonging to the Lie algebra of $U(n)$.

Example 2.30. The Lie algebra of $SU(n)$ denoted by $\mathfrak{su}(n)$ consists of all complex matrices satisfying:

$$\mathfrak{su}(n) = \{A \in \mathcal{M}_n(\mathbb{C}); A^* = -A, \text{trace}(A) = 0\}$$

Indeed, the condition $A^* = -A$ follows from the previous example, and as in the previous results adding the condition of determinant 1 of the special group, it translates to the condition "trace 0" to the algebra.

Example 2.31. The Lie algebra of $O(n)$ denoted by $\mathfrak{o}(n)$ consists of all real matrices satisfying:

$$\mathfrak{o}(n) = \{A \in \mathcal{M}_n(\mathbb{R}); A^{tr} = -A\}$$

Example 2.32. The Lie algebra of $SO(n)$ denoted by $\mathfrak{so}(n)$ consists of all real matrices satisfying:

$$\mathfrak{so}(n) = \{A \in \mathcal{M}_n(\mathbb{R}); A^{tr} = -A\}$$

For the Lie groups $O(n)$ and $SO(n)$ an exactly similar argument follows for finding their Lie algebras.

As we mentioned earlier, a Lie group and its Lie algebra are related by a map as follows:

Definition 2.33. If G is a matrix Lie group with Lie algebra \mathfrak{g} , then the *exponential map* for G is defined by:

$$\begin{aligned} \exp : \mathfrak{g} &\longrightarrow G \\ A &\mapsto \exp(A). \end{aligned}$$

Given the above map, it is natural to ask whether for every element in G , there exists an element in \mathfrak{g} which can be matched by the exponential map. The answer turns out to be negative, as shown by the following example.

Example 2.34. Let $A \in SL(2; \mathbb{C})$ be the matrix given by

$$A = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix}$$

Then there does not exist a matrix $X \in \mathfrak{sl}(2; \mathbb{C})$ such that $\exp(X) = A$. Indeed, if $X \in \mathfrak{sl}(2; \mathbb{C})$ has distinct eigenvalues, then X is diagonalizable and $\exp(X)$ too, unlike the matrix A . If $X \in \mathfrak{sl}(2; \mathbb{C})$ has a repeated eigenvalue, this eigenvalue must be 0 or the trace of X would not be zero. Thus, there is a nonzero vector v with $Xv = 0$. Then $e^X v = e^0 v = v$. We conclude that the eigenvalue of e^X is 1, unlike the matrix A .

2.4 Matrix Norm

The proof of the Solovay-Kitaev Theorem relies on the fact that we can equip the group $SU(2)$ with a norm and induce a distance between all the elements of the group. As we will see in Chapter 3 and 4 we will measure how close are two elements of $SU(2)$ using this distance. In order to prove the Solovay-Kitaev Theorem the most common norm used is the trace norm, however since $SU(2)$ is a finite dimensional space, all norms can be related to each other by constant factors.

The trace norm is defined as:

Definition 2.35. Let $A \in \mathcal{M}_n(\mathbb{C})$. The *trace norm* of A is defined as:

$$\|A\| := \text{trace} \sqrt{A^* A}$$

If A is normal ($A^* A = A A^*$) and $\lambda_1 \dots \lambda_n$ are the eigenvalues of A , this norm can be computed as $\|A\| = \sum_{i=1}^n |\lambda_i|$. The trace norm satisfies the following properties:

1. Triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$.
2. Submultiplicativity: $\|AB\| \leq \|A\| \cdot \|B\|$.
3. Unitary invariance: $\|UAV\| = \|A\|$ for all U, V belonging to $U(n)$.

Let $d(\cdot, \cdot) : \mathcal{M}_n(\mathbb{C}) \rightarrow \mathbb{R}$ be the metric induced by the trace norm, i.e:

$$d(A, B) = \|A - B\|$$

The following property will play a key role in the proof of the Solovay-Kitaev Theorem:

Definition 2.36. A distance in $\mathcal{M}_n(\mathbb{C})$,

$$d(\cdot, \cdot) : \mathcal{M}_n(\mathbb{C}) \rightarrow \mathbb{R},$$

is called *unitary biinvariant* if and only if

$$d(A, B) = d(AU, BU) = d(UA, UB),$$

for any $U \in U(n)$.

Note that the distance d defined above is clearly unitary biinvariant, since the trace norm has the property of unitary invariance.

Remark 2.37. Some texts use a different norm, called the *operator norm*. This norm is defined as:

$$\|A\|_{op} = \sup_{v \neq 0} \frac{|Av|}{|v|}$$

Equivalently, $\|A\|_{op}^2$ is the largest eigenvalue of the operator A^*A .

2.5 Unitary transformations

The unitary groups can be defined in terms of quantities that are left invariant. In the next section on quantum mechanics we will see why such transformations are so useful for physicists.

Consider a complex transformation in two dimensions given by $x' = Ax$, $x, x' \in \mathbb{C}^2$ where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ with } a, b, c, d \in \mathbb{C}. \quad (1)$$

This transformation is *unitary* if and only if $|x|^2$ remains invariant. Equivalently, if $x = (x_1, x_2)$, A is unitary if and only if $|x_1|^2 + |x_2|^2$ remains invariant. This is equivalent to:

$$|a|^2 + |c|^2 = 1 \quad |b|^2 + |d|^2 = 1 \quad ab^* + cd^* = 0$$

Note from Example 2.11 that these conditions are equivalent to asking that the above square matrix is in $U(2)$. So by simple calculations we see that the Lie group $U(n)$ is the group of unitary transformations.

In addition if we want that the determinant of the unitary transformation to be one, all conditions can be summarized in the following way:

$$A = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}, \quad |a|^2 + |b|^2 = 1 \quad (2)$$

That is the form of the elements in $SU(2)$. Note that there are three-free parameters, since a, b are complex numbers, each have two parameters (the real part and the imaginary part). Also defining the map $\varphi : \mathbb{C}^2 \rightarrow SU(2)$ as,

$$\varphi(a, b) = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \text{ with } a, b \in \mathbb{C},$$

it can be viewed that the 3-sphere S^3 is diffeomorphic to $SU(2)$. Consider $a = x + iy$ and $b = z + it$, with $x, y, z, t \in \mathbb{R}$. If $|a|^2 + |b|^2 = 1$ it is equivalent to

$$x^2 + y^2 + z^2 + t^2 = 1.$$

This is the equation of the 3-sphere S^3 . Thus, restricting φ to S^3 , we have that $\varphi : S^3 \rightarrow SU(2)$ is clearly injective and surjective. Furthermore this map is smooth. Hence, as a manifold, S^3 is diffeomorphic to $SU(2)$, and S^3 can be viewed as a Lie group.

Remark 2.38. In higher dimensions a unitary transformation is a transformation $A : \mathbb{C}^n \rightarrow \mathbb{C}^n$ such that for every $x \in \mathbb{C}^n$, $|Ax| = |x|$. Similarly these transformations can be identified with an element in $U(n)$.

2.6 The Lie algebra of SU(2)

We learned that the exponential map is, in general, not surjective. For our later discussions it will be useful to connect the particular matrix Lie group $SU(2)$ to its Lie algebra. In this section we give some results to do this.

Consider the following three matrices:

$$u_1 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, u_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ and } u_3 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

These matrices belong to $\mathfrak{su}(2)$ and in fact, are generators of $\mathfrak{su}(2)$.

In physics, when dealing with $SU(2)$ or $\mathfrak{su}(2)$ it is usual to work with the *Pauli matrices*, defined as:

$$\sigma_x = -iu_1 \qquad \sigma_y = iu_2 \qquad \sigma_z = -iu_3$$

The Pauli matrices are extensively used in quantum mechanics to represent the spin of fundamental particles. In fact, they are introduced because physicists are used to work with Hermitian operators in quantum mechanics, and multiplying by the factor i the generators of $\mathfrak{su}(2)$ they become Hermitian ($\sigma_i^* = \sigma_i$). In addition, they are unitary matrices.

This set of matrices satisfies the commutation and anticommutation relations:

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \qquad \{\sigma_i, \sigma_j\} = 2\delta_{ij}I$$

Where ϵ_{ijk} denotes the *Levi-Civita symbol* and δ_{ij} is the Kronecker delta, both defined as:

$$\epsilon_{ijk} = \begin{cases} 1, & \text{if } (i, j, k) \text{ is } (1, 2, 3), (2, 3, 1), \text{ or } (3, 1, 2) \\ -1, & \text{if } (i, j, k) \text{ is } (3, 2, 1), (1, 3, 2), \text{ or } (2, 1, 3) \\ 0, & \text{if } i = j, i = k, \text{ or } j = k \end{cases}$$

$$\delta_{ij} = \begin{cases} 1, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}$$

Sometimes it will be useful to consider *the Pauli vector*, defined as:

$$\sigma = (\sigma_x, \sigma_y, \sigma_z)$$

Given a point $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ we will denote the map $u : \mathbb{R}^3 \rightarrow SU(2)$ as

$$u(x) := \exp\left(-\frac{i}{2}x \cdot \sigma\right), \tag{3}$$

where $x \cdot \sigma$ denotes

$$x \cdot \sigma = x_1\sigma_x + x_2\sigma_y + x_3\sigma_z.$$

Note from the definitions of σ_i , that $i\sigma_i \in \mathfrak{su}(2)$ and using Theorem 2.25 it is clear that $ix \cdot \sigma \in \mathfrak{su}(2)$.

Let x be decomposed as

$$x = \varphi \cdot n = \varphi \cdot (n_x, n_y, n_z)$$

such that $\varphi \in \mathbb{R}$ and n is a unit vector. Then if we compute $u(x)$ simply by taking Definition 2.15 and making the calculations, we obtain:

$$u(x) = \begin{pmatrix} \cos(\frac{1}{2}\varphi) - in_z \sin(\frac{1}{2}\varphi) & -(n_y + in_x) \sin(\frac{1}{2}\varphi) \\ (n_y - in_x) \sin(\frac{1}{2}\varphi) & \cos(\frac{1}{2}\varphi) + in_z \sin(\frac{1}{2}\varphi) \end{pmatrix}.$$

One may easily verify that this is in the form of Equation 2. Thus we have

Lemma 2.39. *For any element U in $SU(2)$, we can choose $x \in \mathbb{R}^3$ such that $u(x) = U$.*

This is why it is said that the Pauli matrices are *infinitesimal generators* of $SU(2)$.

2.7 Connecting SU(2) with SO(3)

In the last part of this section we are going to see the relation between the Lie groups $SU(2)$ and $SO(3)$.

The groups $SU(2)$ and $SO(3)$ are not isomorphic, but one can define a group homomorphism $\Phi : SU(2) \rightarrow SO(3)$ that is *two-to-one* and onto, leading to an isomorphism

$$SU(2)/\mathbb{Z}_2 \cong SO(3).$$

Consider the space V of all 2×2 complex matrices which are self-adjoint and have trace zero. Elements of V are the matrices X of the form

$$X = \begin{pmatrix} x_1 & x_2 + ix_3 \\ x_2 - ix_3 & -x_1 \end{pmatrix}$$

with $x_1, x_2, x_3 \in \mathbb{R}$. If we identify V with \mathbb{R}^3 by the coordinates x_1, x_2, x_3 the standard inner product on \mathbb{R}^3 can be computed as:

$$\langle X, X' \rangle = \frac{1}{2} \text{trace}(X, X') = x_1 x'_1 + x_2 x'_2 + x_3 x'_3.$$

For each $U \in SU(2)$, define the map $\Phi_U : V \rightarrow V$ by:

$$\Phi_U(X) = UXU^{-1}$$

Since U is unitary, we have

$$\Phi_U(A) = (UAU^{-1})^* = UAU^{-1},$$

showing that $\Phi_U(A)$ is again in V . It is easy to check that for all $U_1, U_2 \in U(n)$

$$\Phi_{U_1 U_2} = \Phi_{U_1} \Phi_{U_2}.$$

Furthermore, each Φ_U preserves the inner product $\text{trace}(X, X')/2$:

$$\frac{1}{2}\text{trace}((UXU^{-1})(UX'U^{-1})) = \frac{1}{2}\text{trace}(UXX'U^{-1}) = \frac{1}{2}\text{trace}(XX')$$

where we have used that the trace is invariant under conjugation. It follows that the map $U \rightarrow \Phi_U$ is a homomorphism of $SU(2)$ into the group of orthogonal linear transformations of $V \cong \mathbb{R}^3$, that is $O(3)$ (see Example 2.13). Note that $\Phi_U = \Phi_{-U}$ for any $U \in SU(2)$, so, the map Φ_U is not one to one. In order to prove that Φ_U actually lies in $SO(3)$ and that is a two to one map, we will use the following proposition, whose proof can be found in [Hal15, Page 24].

Proposition 2.40. *The map $U \rightarrow \Phi_U$ is a 2 – 1 and onto map of $SU(2)$ to $SO(3)$ with kernel equal to $\{I, -I\}$.*

3 Quantum computing

In this section we present the qubit, which is the basic unit of information in quantum computing. In order to understand the qubit we will use the Bloch sphere, a geometrical representation of the qubit that allows to see in a clear way the differences with the classical bit. Then, we introduce the concepts of a quantum gate, which plays the role of the operations that we can perform in a quantum computer. Quantum gates are the analogues of the logic gates in classical computing, but they are very different since the operations that we can perform in a quantum system are unitary operations from the group $U(n)$. Finally we give the notion of a universal set of quantum gates, which are a set of unitary operations such that they can generate an arbitrary unitary operation with a given error. The discussion of the optimality of these sets will be given in the next section, connecting with the main result of this work: The Solovay-Kitaev theorem.

3.1 Qubit

Popular science has introduced the notion of a qubit as a classical bit being in the superposition of 0 and 1 at the same time. This point of view is the easiest way to understand what is a qubit, but in general a qubit is much more than that. As this is a mathematics text, we pretend to give a more rigorous definition and hopefully a much richer vision.

First of all, let us briefly present the bra-ket notation, widely used in quantum mechanics. Let $\psi = (a_1, \dots, a_n) \in \mathbb{C}^n$ be a point in \mathbb{C}^n . The *ket* will be defined simply as $\psi = |\psi\rangle$ and we will represent it as a column vector. The *bra*, will be understood as the adjoint of the ket, denoted $\langle\psi|$ and represented by a row vector:

$$|\psi\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \langle\psi| = (a_1^* \quad a_2^* \quad \dots \quad a_n^*)$$

A *braket* will be just the product between the bra and the ket understood as:

$$\langle\psi|\psi\rangle = a_0 a_0^* + a_1 a_1^* + \dots + a_n a_n^*$$

For physicists, a *n-quantum level system* is just a point $|\psi\rangle$ in \mathbb{C}^n . The space of all these points forms a complex Hilbert space. We briefly recall that a complex *Hilbert space* is a complex vector space \mathcal{H} , equipped with an inner product, such that the norm turns \mathcal{H} into a complete metric space. Given $|\psi\rangle, |\varphi\rangle \in \mathcal{H}$, the inner product is defined as the bra-ket between the two states:

$$\langle\varphi|\psi\rangle = b_1^* a_1 + \dots + b_n^* a_n \text{ where } a_0, \dots, a_n \text{ and } b_0, \dots, b_n$$

are the coefficients of the vectors $|\psi\rangle$ and $|\varphi\rangle$.

Physically, a vector $|\psi\rangle$ represents a superposition of n -classical states, and this classical states are just the vector basis where $|\psi\rangle$ is defined. For example,

given an electron, with two spin states up $|\uparrow\rangle$ and down $|\downarrow\rangle$. We will represent the spin states of the electron as $|\psi\rangle = a_1 |\uparrow\rangle + a_2 |\downarrow\rangle$, and if we write $|\psi\rangle$ in the basis

$$\{|\uparrow\rangle, |\downarrow\rangle\}, |\psi\rangle = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}.$$

Actually the coefficients a_1, a_2 can be understood as the probabilities of the electron to be with spin up or with spin down. In quantum mechanics these probabilities are computed with the inner product of two vectors, so since the electron can be only with spin up or down it is natural to give the vector normalized: $|a_0|^2 + |a_1|^2 = 1$. For that reason the next definition arises:

Definition 3.1. A *pure state* of a n -quantum level system is a vector $|\psi\rangle$ in \mathbb{C}^n such that $\langle\psi|\psi\rangle = 1$.

Moreover, in quantum mechanics it is not possible to distinguish physical states that are the same up to a global phase factor, it means that we can not distinguish between $|\psi\rangle$ and $e^{i\theta} |\psi\rangle$. For that reason it is convenient to introduce the following equivalence relation:

$$\psi \sim \psi' \iff \psi = e^{i\theta} \psi'$$

for some $\theta \in \mathbb{R}$. This equivalence relation leads naturally to think about the equivalence classes as "complex" lines through the origin in \mathbb{C}^n . These lines form the *complex projective space*

$$\mathbb{C}\mathbb{P}^{n-1} := \frac{\mathbb{C}^n - \{0\}}{z \sim \lambda z}, \lambda \in \mathbb{C}^*.$$

There is a one to one correspondence between points in $\mathbb{C}\mathbb{P}^{n-1}$ and n -quantum level systems. With this identification we are ready to give the definition of a qubit.

Definition 3.2. A *qubit* is a 2-level quantum system understood as a complex projective line $\mathbb{C}\mathbb{P}^1$. A *pure qubit state* is a point in $\mathbb{C}\mathbb{P}^1$.

Remark 3.3. In general the standard basis used for a qubit is $\mathcal{B} = \{|0\rangle, |1\rangle\}$. The 0 and 1 are used to make an analogy between the classical bits that can be 0 or 1. Nowadays, there is not a clear quantum system to make a qubit but for example the $|0\rangle$ and $|1\rangle$ can be physically realized by photons and assign these states to states of polarization that can be "horizontal" or "vertical". The physical realization of the qubit is one of the most active fields in physics now since there are lot of ways to implement them.

Since there is a correspondence between geometry and the definition of a qubit, we next describe a standard way to visualize geometrically the qubit.

3.2 The density matrix

In quantum mechanics there are other physical states apart of the pure states which are called *mixed states*. These states can be understood as an *ensemble of pure states*. More precisely, given n pure states $|\psi_i\rangle$, $1 \leq i \leq n$, let p_i be the probability of the quantum system to be in the state $|\psi_i\rangle$. Then a *mixed state* is defined by

$$\{p_i, |\psi_i\rangle\}.$$

To describe these states we will use *the density matrix* defined by the following equation:

$$\rho \equiv \sum_i p_i |\psi_i\rangle \otimes \langle \psi_i|$$

Note that the density matrix of a pure state $|\psi\rangle$ is $\rho = |\psi\rangle \otimes \langle \psi|$. This alternate formulation is equivalent to the state vector approach (vectors in \mathbb{C}^n), but in quantum mechanics there are lot of scenarios which is more useful to work with the density matrix. For us, the density matrix will be useful because it provides an easy way of visualizing geometrically the qubit with *the Bloch Sphere*.

3.3 Bloch Sphere

The Bloch sphere is the most common representation of points in \mathbb{CP}^1 used by Physicists. We briefly explain how to relate points in \mathbb{CP}^1 and points in a sphere S^2 .

Without loss of generality we can write the state of a pure qubit as:

$$|\psi\rangle = \begin{pmatrix} \cos(\frac{\theta}{2}) \\ e^{i\phi} \sin(\frac{\theta}{2}) \end{pmatrix} \quad (4)$$

with $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$. The factor $\frac{1}{2}$ is introduced to arrange the intervals like the spherical coordinates. With this convention, we can visualize the space of qubits in Figure 1.

The states of the basis $|0\rangle$ and $|1\rangle$ are placed at the north pole and the south pole of the sphere respectively. This suggest that there must be a one to one correspondence between pure qubit states and the points of a unit sphere S^2 in \mathbb{R}^3 . Indeed, a pure qubit state was defined as a point of \mathbb{CP}^1 which is homeomorphic to the sphere S^2 . To see that define a *Bloch vector* r in \mathbb{R}^3 as $r = (x, y, z)$:

$$\begin{cases} x = \sin(\theta) \cos(\phi) \\ y = \sin(\theta) \sin(\phi) \\ z = \cos(\theta) \end{cases}$$

The density matrix of 4 can be written as:

$$\rho = |\psi\rangle \otimes \langle \psi| = \frac{1}{2} \begin{pmatrix} 1 + \cos(\theta) & e^{-i\phi} \sin(\theta) \\ e^{i\phi} \sin(\theta) & 1 - \cos(\theta) \end{pmatrix} = \frac{1}{2}(I + x\sigma_x + y\sigma_y + z\sigma_z)$$

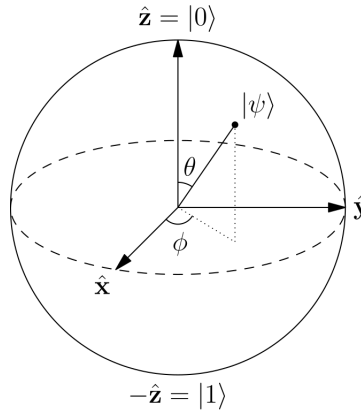


Figure 1: Bloch sphere

If $r = (x, y, z)$ and σ denotes the the Pauli vector we can rewrite the density matrix as:

$$\rho = \frac{1}{2}(I + r \cdot \sigma)$$

This way of writing the density matrix relates pure state qubits with points in the unit sphere S^2 in \mathbb{R}^3 . Moreover, we can relate the inner products for the pure qubit states and the standard inner product in \mathbb{R}^3 . If r_1 and r_2 are the Bloch vectors for the pure qubit states $|\psi_1\rangle$ and $|\psi_2\rangle$. Then we have:

$$|\langle \psi_1 | \psi_2 \rangle|^2 = \text{trace}(\rho_1 \rho_2) = \frac{1}{2}(1 + r_1 \cdot r_2).$$

Remark 3.4. If we have two pure qubit states that are orthogonal, $(\langle \varphi | \varphi' \rangle) = 0$, from the last relation we see that $r_1 \cdot r_2 = -1$. So in the Bloch sphere picture orthogonal states correspond to antipodal points.

Remark 3.5. If we work with mixed states, the same construction gives that the points inside the Bloch sphere corresponds to mixed states. Thus, the surface of the Bloch sphere correspond to pure qubit states, and the inside to mixed states.

3.4 Single qubit operations

Given the basic notions of the interpretation of the qubit it is time to describe how we operate on them. Operations on a qubit must preserve the inner product with itself in order to maintain the probabilities for the states. As we saw in Section 1.5 this type of transformations are the 2×2 unitary matrices. Remember that a pure state qubit is represented by a point $z = (z_0, z_1) \in \mathbb{C}P^1$, so a unitary transformation $U \in U(2)$ will act as a matrix vector product, $U \cdot x$. Then, we can provide a definition analogous to the classical logical gate in the quantum case.

Definition 3.6. A *quantum gate* for a pure state qubit is a unitary transformation $U : \mathbb{CP}^1 \rightarrow \mathbb{CP}^1$.

There are some important quantum gates that will be useful to consider. Of these, some of the most important are the Pauli matrices:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note that when we talk about Pauli matrices in terms of quantum gates we change the notation. Three other quantum gates that play a huge role in quantum computation are the Hadamard gate (denoted H), phase gate (denoted S), and $\pi/8$ gate (denoted T):

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}; \quad S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}; \quad T = \begin{pmatrix} 1 & 0 \\ 0 & \exp(i\pi/4) \end{pmatrix}.$$

As we will see in Section 3.5, H and T are a universal set of quantum gates or equivalently, they generate a dense subset in $U(2)$.

With the picture of the Bloch sphere we can try to visualize this type of operations with rotations. Since a qubit is a point of the Bloch sphere, a quantum gate sends this point to another, so intuitively one can think of these quantum gates as rotations over some axis. In fact, when we exponentiate the Pauli matrices we find the *rotation operators*:

$$R_x(\theta) \equiv e^{-i\theta X/2} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} X = \begin{pmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix} \quad (5)$$

$$R_y(\theta) \equiv e^{-i\theta Y/2} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} Y = \begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix} \quad (6)$$

$$R_z(\theta) \equiv e^{-i\theta Z/2} = \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} Z = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix} \quad (7)$$

The calculations for $R_i(\theta)$ can be done easily using proposition 2.19.

If you take a general qubit state (4) and apply one of these operators it follows that they are rotations about the x, y, z axes of the Bloch sphere.

Example 3.7. Consider the following state:

$$|\psi\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \cos(\pi/4) \\ \sin(\pi/4) \end{pmatrix}$$

From (4) we see that the Bloch sphere angles θ, ϕ correspond to $\theta = \frac{\pi}{2}$ and $\phi = 0$. If we apply a rotation over the y axis of $\theta = \pi/2$ to the state $|\psi\rangle$ it will be brought to the south pole corresponding to the state $|1\rangle$. Let's see:

$$R_y\left(\frac{\pi}{2}\right) |\psi\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = |1\rangle.$$

In general it can be shown that a rotation of an angle θ about a general axis $n = (n_x, n_y, n_z)$ (normalized to 1) of the Bloch sphere can be expressed as:

$$R_n(\theta) \equiv \exp(-i\theta n \cdot \sigma/2) = \cos(\theta/2)I - i \sin(\theta/2)(n_x X + n_y Y + n_z Z). \quad (8)$$

Example 3.8. The Hadamard gate in the Bloch sphere can be expressed as a product of rotations over the x and z axis up to a global phase. By direct calculation:

$$H = e^{i\frac{\pi}{2}} R_z\left(\frac{\pi}{2}\right) R_x\left(\frac{\pi}{2}\right) R_z\left(\frac{\pi}{2}\right)$$

Example 3.9. The $\pi/8$ gate T is, up to a global phase a rotation by $\pi/4$ radians around the z axis on the Bloch sphere.

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix} = e^{i\frac{\pi}{8}} \begin{pmatrix} e^{-i\frac{\pi}{8}} & 0 \\ 0 & e^{i\frac{\pi}{8}} \end{pmatrix} = e^{i\frac{\pi}{8}} R_z\left(\frac{\pi}{4}\right).$$

Example 3.10. The product of Hadamard gates and $\pi/8$ gates HTH is up to a global phase a rotation by $\pi/4$ radians around the x axis on the Bloch Sphere. By direct calculation:

$$R_x\left(\frac{\pi}{4}\right) = e^{-i\frac{\pi}{8}} HTH.$$

With these results one can think if is it possible to express any unitary operator as a composition of rotations. Indeed, we have the following result:

Proposition 3.11. *Suppose that $U \in U(2)$ is a quantum gate. Then there exist real numbers α, β, γ and δ such that:*

$$U = e^{i\alpha} R_z(\beta) R_y(\gamma) R_z(\delta).$$

Proof. From the definitions of the rotation matrices it follows that:

$$e^{i\alpha} R_z(\beta) R_y(\gamma) R_z(\delta) = \begin{pmatrix} e^{i(\alpha-\beta/2-\delta/2)} \cos\frac{\gamma}{2} & -e^{i(\alpha-\beta/2+\delta/2)} \sin\frac{\gamma}{2} \\ e^{i(\alpha+\beta/2-\delta/2)} \sin\frac{\gamma}{2} & e^{i(\alpha+\beta/2+\delta/2)} \cos\frac{\gamma}{2} \end{pmatrix}$$

From Section 2.5, imposing that U is unitary it can be shown that the general expression coincides with the calculation above. \square

Actually there exists other forms of expressing a 2×2 unitary matrix with the product of rotation operators. Remember from Section 1.7 that there was a map between $SU(2)$ and $SO(3)$, so it is quite natural that unitary operators can be viewed as rotations. In particular the above theorem can be generalized with the following result:

Proposition 3.12. *Let m and n be non-parallel real unit vectors of \mathbb{R}^3 . For any quantum gate U acting on a single qubit there exist real numbers α, β, γ and δ such that:*

$$U = e^{i\alpha} R_n(\beta) R_m(\gamma) R_n(\delta).$$

3.5 Universal quantum gates

At this point one may wonder why we have focused in studying only a single qubit. As anyone can imagine a single qubit is not useful for real computations. For example, a physical system with two qubits can simulate the classical states 00, 01, 10 and 11. Moreover this system can actually be a superposition of all these states, so there's an exponential gain in terms of computational power. If we work with a system of n -qubits we would be able to simulate a superposition of 2^n classical states.

From our definition of quantum gate, it can be generalized to systems to n -qubits as unitary matrices of dimension $2^n \times 2^n$. At this point it seems that we would need to study operations in higher dimensions. What happens is that any unitary operation on n qubits can be decomposed as operations on a single qubit and a extra gate called controlled not (CNOT) in $U(4)$. The gate CNOT acts on 2 qubits. In the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ the gate CNOT acts flipping the second qubit if the first qubit is 0, which means:

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\text{CNOT} |00\rangle = |00\rangle, \text{CNOT} |01\rangle = |01\rangle, \text{CNOT} |10\rangle = |11\rangle, \text{CNOT} |11\rangle = |10\rangle.$$

We refer to Section 4 of [NC00] for a proof of this fact. For this reason, it is important to study the single qubit operations.

In practice, if one wishes to compute with a quantum computer, the set of quantum operations will be finite. It is important to remark that all these unitary transformations will translate to physical operations in some system, so we will be restricted. Furthermore, these physical operations seems to be very difficult to control in the laboratories, so in principle we will have access to a small number of operations. This is why it is useful to find what we call a *set of universal quantum gates*. Let's introduce some basic definitions to introduce this concept.

Definition 3.13. Let $d, \ell \geq 1$ be integers and \mathcal{G} a finite set of elements of $SU(2)$. A *word of length ℓ of elements in \mathcal{G}* is given by $w_\ell = g_1 g_2 \dots g_\ell$, with $g_i \in \mathcal{G}$. Denote by \mathcal{G}_ℓ the set of all words in \mathcal{G} of length at most ℓ : and by $\langle \mathcal{G} \rangle$ the set of all words in \mathcal{G} of finite length:

$$\mathcal{G}_\ell = \{w_k = g_1 g_2 \dots g_k; g_i \in \mathcal{G}, k \leq \ell\} \text{ and } \langle \mathcal{G} \rangle = \bigcup_{\ell < \infty} \mathcal{G}_\ell.$$

Remark 3.14. Note that in general, such a word $w_\ell = g_1 g_2 \dots g_\ell \in \mathcal{G}_\ell$ is not necessarily an element of \mathcal{G} .

Remark 3.15. Since matrices in $U(2)$ and $SU(2)$ differ from a constant factor we will give most of the definitions in $SU(2)$.

Informally a set of quantum gates \mathcal{G} is called a *set of universal quantum gates* if for any quantum gate U , it can be expressed as some word w_ℓ from \mathcal{G} . It is not difficult to see that this definition is not very useful. Note that $SU(2)$ has infinite elements, and the number of finite sequences from a finite set is countable. To solve this problem we use the concept of density. Recall we have a notion of distance for matrices introduced in Section 2.4.

Definition 3.16. A set of quantum gates \mathcal{G} is called a *set of universal quantum gates* if $\langle \mathcal{G} \rangle$ is dense in $SU(2)$, i.e., if and only if for every element $U \in SU(2)$ and for all $\epsilon > 0$, there exists $g \in \langle \mathcal{G} \rangle$ such that $d(g, U) < \epsilon$.

From this definition $d(g, U)$ can be understood as the error of implementing g instead of U . More precisely:

Definition 3.17. Let $U, V \in U(2)$. The *error* when V is implemented instead of U is defined by:

$$E(U, V) := \max_{\|\psi\rangle=1} \|(U - V)|\psi\rangle\|.$$

From Section 2.4, this definition of error is equivalent to the distance between the operators U and V induced by the operator norm. Moreover, the error can be defined using another distance like the trace norm. Thus, the concepts of universal set of quantum gates can be viewed as a set that can implement an arbitrary unitary operation with an arbitrary non-zero error.

Proposition 3.18. *Given two sequences V_1, V_2, \dots, V_m U_1, U_2, \dots, U_m of quantum gates, the error of implementing the first sequence instead of the second satisfies:*

$$E(U_m \cdots U_2 U_1, V_m \cdots V_2 V_1) \leq \sum_{i=1}^m E(U_i, V_i).$$

Proof. To prove the result it suffices to prove the first case for $m = 2$. The general case follows immediately by induction.

$$\begin{aligned} E(U_2 U_1, V_2 V_1) &= \max_{\|\psi\rangle=1} \|(U_2 U_1 - V_2 V_1)|\psi\rangle\| \\ &= \max_{\|\psi\rangle=1} \|(U_2 U_1 - V_2 U_1) + (V_2 U_1 - V_2 V_1)|\psi\rangle\| \end{aligned}$$

Using the triangle inequality, and the biinvariance of the distance we get:

$$\begin{aligned} E(U_2 U_1, V_2 V_1) &\leq \max_{\|\psi\rangle=1} \|(U_2 - V_2)U_1|\psi\rangle\| + \max_{\|\psi\rangle=1} \|(U_1 - V_1)U_2|\psi\rangle\| \\ &= E(U_2, V_2) + E(U_1, V_1) \end{aligned}$$

□

The following result can be found in Chapter 4 of [NC00]. We give a proof for completeness.

Theorem 3.19. *The Hadamard gate H and the $\pi/8$ gate T are a set of universal quantum gates for $SU(2)$.*

Proof. Consider the gates T and HTH . We are going to show that a successive product of these gates can be used to approximate any single qubit gate with an arbitrary accuracy.

From examples 3.9 and 3.10 the product $THTH$ can be expressed as:

$$\begin{aligned} THTH &= R_z\left(\frac{\pi}{4}\right)R_x\left(\frac{\pi}{4}\right) = \left(\cos\frac{\pi}{8}I - i\sin\frac{\pi}{8}Z\right)\left(\cos\frac{\pi}{8}I - i\sin\frac{\pi}{8}X\right) = \\ &= \cos^2\frac{\pi}{8}I - i\left(\cos\frac{\pi}{8}(X+Z) + \sin\frac{\pi}{8}Y\right)\sin\frac{\pi}{8} \end{aligned}$$

where we have used that $Y = -iZX$. From the expression of a general rotation over the Bloch sphere 8, we see that this product corresponds to a rotation of axis

$$n = \left(\cos\frac{\pi}{8}, \sin\frac{\pi}{8}, \cos\frac{\pi}{8}\right)$$

and through an angle θ corresponding to $\cos(\theta/2) = \cos^2\frac{\pi}{8}$. Note that the vector n is not normalized, thus, to see this result one should do some basic manipulations. This θ is an irrational multiple of π . (This result can be found in [BMP+99]) Let $R_n(\theta)$ denote this new rotation.

Iterating this new rotation we will find that we can achieve with a certain error any rotation of angle α around the n axis with a certain accuracy. Let $\delta > 0$ be the desired accuracy, and N an integer larger than $2\pi/\delta$. Let $\theta_k \in [0, 2\pi)$ be the angle achieved after iterating k times $R_n(\theta)$, $\theta_k = (k\theta) \bmod(2\pi)$. Then using the pigeonhole principle it implies that there exists k and j in the interval $1, \dots, N$ such that $|\theta_k - \theta_j| \leq 2\pi/N < \delta$, thus $|\theta_{k-j}| < \delta$. Using that θ is an irrational multiple of 2π and $j \neq k$ it follows that the sequence $\theta_{\ell(k-j)}$ fills up the interval $[0, 2\pi)$ with angles separated by no more than δ .

With that construction we can say that for any $\epsilon > 0$ there exists an n such that:

$$E(R_n(\alpha), R_n(\theta)^n) < \epsilon$$

From Proposition 3.12 we see that if we can approximate a rotation over another axis different from n the result will follow immediately. Simple algebra implies that for any α :

$$HR_n(\alpha)H = R_m(\alpha)$$

where $m = (\cos\frac{\pi}{8}, -\sin\frac{\pi}{8}, \cos\frac{\pi}{8})$. Using the same iterations as before we find that:

$$E(R_m(\alpha), R_m(\theta)^n) < \epsilon$$

Using Proposition 3.12, any quantum gate U action on a single qubit may be written as:

$$U = R_n(\beta)R_m(\gamma)R_n(\delta)$$

up to a global phase. Finally, using Proposition 3.18 and the approximated rotations, there exists positive integers, n_1, n_2, n_3 such that:

$$E(U, R_n(\theta)^{n_1} H R_n(\theta)^{n_2} H R_n(\theta)^{n_3}) < 3\epsilon$$

Note that all these rotations are constructed only using Hadamard gates and $\pi/8$ gates. If we want to approximate any U with an error of δ , use $\epsilon = \delta/3$ and the result follows. \square

From this result it seems that finding a universal set of quantum gates is not difficult. In fact if one wish to prove that a particular set of quantum gates is an universal set it suffices to prove that this set generate some rotations of a irrational multiple of π . Then this rotations will full fill the interval $[0, 2\pi)$ and since any unitary operation can be expressed as product of rotations the result will follow. In particular finding a particular set is not difficult, in [Llo95] was proved that almost any set of two qubit quantum gates is universal.

4 Solovay-Kitaev

In the last section we presented the concept of universal set of quantum gates in order to approximate any unitary operation. However, we did not give a solid argument about how many gates we would need from our set to achieve the desired approximation. As we mentioned before we will be restricted by the physical systems, and controlling a high number of gates does not seem to be realizable, at least in the first steps of quantum computing.

The Solovay-Kitaev Theorem demonstrates that it is possible to approximate any unitary operation to precision ϵ with $\mathcal{O}(\log^c(1/\epsilon))$ gates from a given set, where c is a constant between 2 and 4 that depends on the implementation. The best value of c that we can achieve is 1 [HBC02], but achieving this value seems to be very difficult.

Remark 4.1. The big \mathcal{O} notation is used to classify functions according to their growth rates. In this work we will use the standard \mathcal{O} which is defined as follows:

Let f be a real or complex function and g a real function. Then

$$f(x) = \mathcal{O}(g(x)) \Leftrightarrow \exists x_0, K \in \mathbb{R}, K > 0 \text{ s.t. } |f(x)| \leq Kg(x) \text{ for any } x \geq x_0.$$

Throughout this section we will review the standard construction with $c = 4$ that is presented in [NC00]. In order to make a more detailed construction, we have added all the important computations that were left as exercises in the book, some of them taken from [Har01] and [Ozo09]. In the last part of the section we briefly sketch some modifications to improve the value of c to 3.

4.1 Preliminaries on distances in $SU(2)$

In order to prove the Solovay-Kitaev Theorem, we will need some previous results. Some of these results provide a relation between the distance in $SU(2)$ induced by the trace norm and the usual distance in \mathbb{R}^3 . These relations will be useful to understand in a clear way the construction of the theorem for matrices in $SU(2)$. Once one has understood the construction in dimension 2, the general case follows by a few modifications.

Recall from Section 2.4, that the trace norm of a matrix $A \in \mathcal{M}_n(\mathbb{C})$ is given by

$$\|A\| = \text{trace}\sqrt{A^*A}.$$

Definition 4.2. Given elements $U, V \in SU(d)$, we will denote their *group commutator* by

$$[U, V]_{gp} := UVU^*V^*.$$

Remark 4.3. The group commutator is related to the usual commutator in the sense that if $U, V \in SU(d)$ satisfy $[U, V] = 0$ then we have $[U, V]_{gp} = I$. So, the group commutator is close to the identity if and only if the two matrices almost-commute.

Proposition 4.4. *Let $\epsilon \geq 0$ and let A, B be Hermitian matrices such that $\|A\|, \|B\| \leq \epsilon$. Then for sufficient small ϵ there is a constant c such that*

$$\|e^{-[A,B]} - [e^{-iA}, e^{-iB}]_{gp}\| \leq c\epsilon^3.$$

Proof. Using the hermicity of A and B and Point 2 of Proposition 2.17

$$[e^{-iA}, e^{-iB}]_{gp} = e^{-iA} e^{-iB} e^{iA} e^{iB}$$

and the series expansion for the matrix exponential

$$[e^{-iA}, e^{-iB}]_{gp} = I - [A, B] + \frac{i}{2}[A + B, A^2 + 2AB + B^2] + \mathcal{O}(\epsilon^4)$$

Computing the other term:

$$e^{-[A,B]} = I - [A, B] + \mathcal{O}(\epsilon^4)$$

Note that all terms are equal up to second order, thus using the triangle inequality and the submultiplicativity of the norm the result follows. \square

In Section 4.4 this result is proved for the case $n = 2$ giving a value for the constant c .

In order to relate the distances between SU(2) and \mathbb{R}^3 , we will use the map $u : \mathbb{R}^3 \rightarrow \text{SU}(2)$ introduced in Section 2.6:

$$u(x) = \exp\left(-\frac{i}{2}x \cdot \sigma\right)$$

The following is a direct computation, using that $[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}$.

Proposition 4.5. *Let $y, z \in \mathbb{R}^3$. Then*

$$\exp\left(-\left[\frac{1}{2}y \cdot \sigma, \frac{1}{2}z \cdot \sigma\right]\right) = u(y \times z).$$

Proposition 4.6. *If $\vec{r} \in \mathbb{R}^3$ then $\|u(\vec{r}) - I\| = 4\sin\frac{|\vec{r}|}{4}$.*

Proof. The eigenvalues of $r \cdot \sigma$ are $\pm |r|$. If $r = (r_x, r_y, r_z)$:

$$\det((r \cdot \sigma) - \lambda I) = \begin{vmatrix} r_z - \lambda & r_x - ir_y \\ r_x + ir_y & -r_z - \lambda \end{vmatrix} = \lambda^2 - r_x^2 - r_y^2 - r_z^2.$$

Thus the eigenvalues of $u(r) - I$ are $e^{\pm \frac{i}{2}|r|} + 1$. From a simple computation we see that $|e^{\pm \frac{i}{2}|r|} + 1| = 2\sin\frac{r}{4}$. Computing the trace norm 2.35 with the sum of the modulus of the eigenvalues the result follows. \square

Remark 4.7. Note that from the first part of the proof it follows that given any $x \in \mathbb{R}^3$ then $\|x \cdot \sigma\| = 2|x|$.

Given an integer $\epsilon \geq 0$, we will denote by S_ϵ the closed ball of radius ϵ and center the identity matrix in $SU(d)$:

$$S_\epsilon := \{s \in SU(2); d(s, I) \leq \epsilon\}.$$

The following proposition relates the distance of an element of $SU(2)$ to the identity and the length of a vector in \mathbb{R}^3

Proposition 4.8. *Let $r \in \mathbb{R}^3$. If $u(r) \in S_\epsilon$ then $|r| < \epsilon + \mathcal{O}(\epsilon^3)$.*

Proof. From 4.6 we have $\|u(r) - I\| = 4 \sin \frac{|r|}{4} < \epsilon$. The inequality comes from the condition $u(r) \in S_\epsilon$. Thus:

$$|r| < 4 \arcsin \left(\frac{\epsilon}{4} \right).$$

The result follows immediately using the Taylor expansion

$$\arcsin(x) = x + \frac{x^3}{2 \cdot 3} + \dots$$

□

Finally, the following proposition will allow us to relate the distance between two elements in $SU(2)$ and the length of two vectors in \mathbb{R}^3 .

Proposition 4.9. *If $y, z \in \mathbb{R}^3$ with $|y| < \epsilon, |z| < \epsilon$, then*

$$\|u(y) - u(z)\| = |y - z| + \mathcal{O}(\epsilon^3).$$

Proof. Let $Y = \frac{1}{2}\sigma y$ and $Z = \frac{1}{2}\sigma z$. Expanding $u(y)$ and $u(z)$ by the definition of matrix exponential and using $(2Y)^2 = (\sigma y)^2 = y^2 I$:

$$\begin{aligned} A = u(y) - u(z) &= -i(Y - Z) - \frac{Y^2 - Z^2}{2} + i\frac{Y^3 - Z^3}{6} + \mathcal{O}(\epsilon^4) \\ &= -i(Y - Z) - \frac{y^2 - z^2}{8} + i\frac{y^2 Y - z^2 Z}{24} + \mathcal{O}(\epsilon^4) \end{aligned}$$

Thus:

$$\begin{aligned} AA^* &= (Y - Z)^2 - \frac{(Y - Z)(y^2 Y - z^2 Z)}{24} - \frac{(y^2 Y - z^2 Z)(Y - Z)}{24} \\ &\quad + \left(\frac{y^2 - z^2}{8} \right)^2 I + \mathcal{O}(\epsilon^5) \\ &= \left(\frac{1}{4}|y - z|^2 + \beta \right) I + \mathcal{O}(\epsilon^2) \end{aligned}$$

where β is defined as:

$$\beta = \frac{y^4 + z^4}{3 \cdot 2^4} - \frac{(y^2 + z^2)y \cdot z}{3 \cdot 2^4} + \left(\frac{y^2 - z^2}{8} \right)^2$$

To do this rearrangement we have used the associativity of the scalar product and $(\sigma_i)^2 = I$. Let $\alpha = |y - x|$, then $\sqrt{AA^*} = \sqrt{\frac{1}{4}(\alpha^2 + \beta)I + \mathcal{O}(\epsilon^5)}$. Since we have expressed $\sqrt{AA^*}$ in terms of the identity matrix, it will be easy to compute the trace norm 2.35 :

$$\|u(y) - u(z)\| = \text{trace}\sqrt{AA^*} = \alpha\sqrt{1 + 4\beta/\alpha^2} + \mathcal{O}(\epsilon^3)$$

Since $\beta \leq 0$, $\|u(y) - u(z)\| \leq \alpha + k\epsilon^3$ for some constant k . □

4.2 Shrinking Lemma

The standard proof of the Solovay-Kitaev Theorem relies on the Shrinking Lemma. This lemma allows us to construct better approximations of unitary matrices around a neighbourhood of the identity at the cost of requiring higher number of generators.

Definition 4.10. Let $\epsilon, \epsilon' \geq 0$ be integers and let S, W be subsets of $SU(2)$. Then:

- (1) The set W is called ϵ -net for S if for all $s \in S$ there exists $w \in W$ such that $d(s, w) \leq \epsilon$, so that S is contained in the ϵ -neighbourhood of W .
- (2) The set W is called (ϵ, ϵ') -net if it is ϵ' -net for S_ϵ , so that the closed ball S_ϵ is contained in the ϵ' -neighbourhood of W .

Given a set of elements \mathcal{G} in $SU(2)$, recall the definitions of $\langle \mathcal{G} \rangle$ and \mathcal{G}_ℓ established in Definition 3.13.

Lemma 4.11 (Shrinking Lemma). *Let \mathcal{G} be a finite set of elements in $SU(2)$ containing its own inverses such that $\langle \mathcal{G} \rangle$ is dense in $SU(2)$. There exist constants $C, \epsilon' > 0$ with $C\epsilon' < 1$ such that, for every $\epsilon \leq \epsilon'$:*

$$\text{If } \mathcal{G}_\ell \text{ is an } (\epsilon, \epsilon^2)\text{-net, then } \mathcal{G}_{5\ell} \text{ is a } (\sqrt{C}\epsilon^{\frac{3}{2}}, C\epsilon^3)\text{-net.}$$

Moreover, the constant ϵ' is independent of \mathcal{G} .

The main idea in the proof of the Lemma is taking group commutators of elements in S_ϵ and proving that these commutators fills S_{ϵ^2} much more densely. The proof is completed using the biinvariance of the distance, which allows to apply a translation step in order to get good approximations for any element of $S_{\sqrt{C}\epsilon^{\frac{3}{2}}}$.

Proof. Assume that \mathcal{G}_ℓ is an (ϵ, ϵ^2) -net, for some $\epsilon > 0$. We first prove that there is a constants C such that $\mathcal{G}_{4\ell}$ is a $(\epsilon^2, C\epsilon^3)$ -net. Let $U \in S_{\epsilon^2}$. Since $U \in SU(2)$, by Section 2.6 we can find $x \in \mathbb{R}^3$ such that

$$U = u(x) = \exp\left(-\frac{i}{2}x \cdot \sigma\right).$$

Combining Proposition 4.6 with $\|u(x) - I\| < \epsilon^2$ and expanding $\sin(x)$ with the Taylor expansion we get:

$$|x| < \epsilon^2 + \mathcal{O}(\epsilon^6)$$

Now choose $y, z \in \mathbb{R}^3$ such that $x = y \times z$ and $|y|, |z| < \epsilon$. Recall that by Proposition 4.6 and using the Taylor expansion of $\sin(x)$ one easily verifies that $u(y), u(z) \in S_\epsilon$.

As \mathcal{G}_ℓ is a (ϵ, ϵ^2) -net, there exist $Y, Z \in \text{SU}(2)$ such that $Y, Z \in \mathcal{G}_\ell \cap S_\epsilon$ and

$$\begin{aligned} \|Y - u(y)\| &\leq \epsilon^2 \\ \|Z - u(z)\| &\leq \epsilon^2. \end{aligned}$$

By Section 2.6 we can find $y_0, z_0 \in \mathbb{R}^3$ such that $Y = u(y_0)$ and $Z = u(z_0)$. From Propositions 4.8 and 4.9 we obtain the following inequalities:

$$\begin{aligned} \|u(y_0) - I\| &< \epsilon \\ \|u(z_0) - I\| &< \epsilon \\ |y_0|, |z_0| &< \epsilon + \mathcal{O}(\epsilon^3) \\ |y_0 - y| &< \epsilon^2, |z_0 - z| < \epsilon^2 \end{aligned}$$

We will now prove that:

$$\|U - [u(y_0), u(z_0)]_{gp}\| < C\epsilon^3$$

To perform this calculation it suffices to use the triangle inequality together with $U = u(y \times z)$. We get:

$$\|U - [u(y_0), u(z_0)]_{gp}\| \leq \|u(y \times z) - u(y_0 \times z_0)\| + \|u(y_0 \times z_0) - [u(y_0), u(z_0)]_{gp}\|.$$

For the first term we use Proposition 4.9:

$$\begin{aligned} \|u(y \times z) - u(y_0 \times z_0)\| &= |y \times z - y_0 \times z_0| + \mathcal{O}(\epsilon^6) = \\ &= |(y - y_0) \times (z - z_0) + y_0 \times (z - z_0) + (y - y_0) \times z_0| + \mathcal{O}(\epsilon^6) \leq \\ &\leq |y - y_0||z - z_0| + |y_0||z - z_0| + |y - y_0||z_0| + \mathcal{O}(\epsilon^6) \\ &\leq 2\epsilon^3 + \mathcal{O}(\epsilon^4) \end{aligned}$$

For the second term we use Propositions 4.4 and 4.5:

$$\begin{aligned} \|u(y_0 \times z_0) - [u(y_0), u(z_0)]_{gp}\| &\leq \\ &\leq \|\exp(-[\frac{1}{2}y_0 \cdot \sigma, \frac{1}{2}z_0 \cdot \sigma]) - [\exp(-\frac{i}{2}y_0 \cdot \sigma), \exp(-\frac{i}{2}z_0 \cdot \sigma)]_{gp}\| \leq c\epsilon^3 \end{aligned}$$

In order to apply Proposition 4.4 note that σ_i are Hermitian and $\|\frac{1}{2}y_0 \cdot \sigma\| < \epsilon$ and $\|\frac{1}{2}z_0 \cdot \sigma\| < \epsilon$. Defining $C = c + 2$ the calculation is done.

Since $u(y_0) = Y \in \mathcal{G}_\ell$ and $u(z_0) = Z \in \mathcal{G}_\ell$, we can conclude that any $U \in S_{\epsilon^2}$ can be approximated with a sequence of 4ℓ elements of \mathcal{G} . This proves that $\mathcal{G}_{4\ell}$ is $(\epsilon^2, C\epsilon^3)$ -net.

Finally we prove that there is a constant ϵ' such that for any $\epsilon \leq \epsilon'$ $\mathcal{G}_{5\ell}$ is a $(\sqrt{C}\epsilon^{\frac{3}{2}}, C\epsilon^3)$ -net.

Now let $U \in S_{\sqrt{C\epsilon^3}}$. The condition that defines ϵ' is that $S_{\sqrt{C\epsilon^3}} \subseteq S_{\epsilon'}$, or equivalently $C\epsilon' < 1$. Since \mathcal{G}_ℓ is a (ϵ, ϵ^2) -net, it follows that for any $\epsilon \leq \epsilon'$ we can find $V \in \mathcal{G}_\ell$ such that:

$$\|U - V\| = \|UV^* - I\| < \epsilon^2,$$

where in the above identity we used the biinvariance of the distance. This result shows that $UV^* \in S_{\epsilon^2}$. Since $\mathcal{G}_{4\ell}$ is $(\epsilon^2, C\epsilon^3)$ -net we can find $y_0, z_0 \in \mathbb{R}^3$ such that

$$\|UV^* - YZY^*Z^*\| < C\epsilon^3,$$

where we defined $Y := u(y_0)$ and $Z := u(z_0)$. Using again the biinvariance of the distance the final result follows:

$$\|UV^* - YZY^*Z^*\| = \|U - YZY^*Z^*V\| < C\epsilon^3 \quad (9)$$

Since $Y, Z, V \in \mathcal{G}_\ell$ it implies $YZY^*Z^*V \in \mathcal{G}_{5\ell}$, thus we can conclude that $\mathcal{G}_{5\ell}$ is a $(\sqrt{C}\epsilon^{\frac{3}{2}}, C\epsilon^3)$ -net. \square

Note that the value of C depends uniquely in the value of c given by Proposition 4.4. This value of C plays a crucial role in the proof of the Solovay-Kitaev Theorem and its important also in the Shrinking Lemma since it determines the condition of the constant ϵ' . As we commented before, this value will be discussed in Section 4.4.

4.3 The Solovay Kitaev Theorem

The Solovay-Kitaev Theorem is one of the most important results in quantum computation since it tells us that any quantum gate can be approximated efficiently given a universal set of quantum gates. Moreover, if we have a sequence of m quantum gates acting on a single qubit, the Solovay Kitaev theorem shows that we can achieve a good approximation for the whole sequence using $\mathcal{O}(m \log^c(m/\epsilon))$ gates from the given gate set, where ϵ denotes the total accuracy of the approximation. A sequence of quantum gates is called a *quantum circuit*, and this result shows that is possible to approximate a quantum circuit with a polylogarithmic increase over the original size of the circuit, which is acceptable for real applications.

With the Shrinking Lemma established we have the necessary tools to give a proof of the Solovay-Kitaev Theorem:

Theorem 4.12 (Solovay-Kitaev Theorem). *Let \mathcal{G} be a finite set of elements in $SU(2)$ containing its own inverses, such that $\langle \mathcal{G} \rangle$ is dense in $SU(2)$. Let $\epsilon > 0$ be a fixed integer. Then \mathcal{G}_ℓ is an ϵ -net in $SU(d)$ for*

$$\ell = \mathcal{O}(\log^c(1/\epsilon)) \text{ with } c = \frac{\log 5}{\log 3/2} \approx 4.$$

Proof. The proof is divided in two stages. First stage is called 'Shrinking' and it consists in building a set of nets such that get very dense in a small neighbourhood of the identity. The second stage is called "Zooming in", in this stage we pick the sequence of nets produced in the previous step, and using the biinvariance of the distance we will be able to translate the nets to the desired points of $SU(d)$ that we want to approximate.

Let's begin with the construction of the two stages.

Shrinking. Since $\langle \mathcal{G} \rangle$ is dense in $SU(2)$ we can choose an $\epsilon_0 \leq \epsilon'$ (the ϵ' comes from Shrinking Lemma 4.11) and find a ℓ_0 such that \mathcal{G}_{ℓ_0} is an ϵ_0^2 -net for $SU(2)$. In particular \mathcal{G}_{ℓ_0} is an $(\epsilon_0, \epsilon_0^2)$ -net. Applying the Shrinking Lemma on \mathcal{G}_{ℓ_0} we find that there is a constant $C > 0$ independent of \mathcal{G} , such that $\mathcal{G}_{5\ell_0}$ is a $(C\epsilon^3, \sqrt{C}\epsilon^{\frac{3}{2}})$ -net. Picking $\mathcal{G}_{5\ell_0}$ and applying again the lemma we obtain that $\mathcal{G}_{5^2\ell_0}$ is a $(\sqrt{C}(C\epsilon^{\frac{3}{2}})^3, C(\sqrt{C}\epsilon^{\frac{3}{2}})^3)$ -net. Iterating k times we find a sequence of nets such that:

$$\mathcal{G}_{5^k\ell_0} \text{ is a } (\epsilon(k), \epsilon(k)^2)\text{-net with } \epsilon(k) = \frac{(C\epsilon_0)^{\frac{3}{2}k}}{C}.$$

From the statement of the Shrinking Lemma remember that $C\epsilon' < 1$, and since $\epsilon_0 \leq \epsilon'$, note that as k increases, $\epsilon(k)$ decreases very fast. Moreover, to be able to perform the next stage we must choose ϵ_0 sufficiently small to achieve that $\epsilon(k)^2 < \epsilon(k+1)$.

Zooming in. Let $U \in SU(2)$ be any element. Since \mathcal{G}_{ℓ_0} is an ϵ_0^2 -net for $SU(2)$, we can find $U_0 \in \mathcal{G}_{\ell_0}$ such that $d(U, U_0) < \epsilon(0)^2$. Define $V = UU_0^*$. Provided that U_0 is unitary, the distance is biinvariant and $\epsilon(k)^2 < \epsilon(k+1)$:

$$d(V, I) = d(V, U_0U_0^*) = d(UU_0^*, U_0U_0^*) = d(U, U_0) < \epsilon(0)^2 < \epsilon(1)$$

Thus $V \in S_{\epsilon(1)}$. As $\mathcal{G}_{5\ell_0}$ is a $(\epsilon(1), \epsilon(1)^2)$ -net, we can find $U_1 \in \mathcal{G}_{5\ell_0}$ such that $d(V, U_1) < \epsilon(1)^2$. Using the same techniques as before:

$$d(U, U_1U_0) = d(UU_0^*, U_1U_0U_0^*) = d(UU_0^*, U_1) = d(V, U_1) < \epsilon(1)^2$$

Thus U_1U_0 is an $\epsilon(1)^2$ approximation for U . Now as before, define $V' = UU_0^*U_1^*$:

$$d(V', I) = d(UU_0^*U_1^*, U_1U_0U_0^*U_1^*) = d(U, U_1U_0) < \epsilon(1)^2 < \epsilon(2)$$

Thus $V \in S_{\epsilon(2)}$. As $\mathcal{G}_{5^2\ell_0}$ is a $(\epsilon(2), \epsilon(2)^2)$ -net we can find $U_2 \in \mathcal{G}_{5^2\ell_0}$ such that $d(V', U_2) < \epsilon(2)^2$.

$$d(U, U_2U_1U_0) = d(UU_0^*U_1^*, U_2U_1U_0U_0^*U_1^*) = d(V', U_2) < \epsilon(2)^2$$

So $U_2U_1U_0$ is an $\epsilon(2)^2$ approximation for U . Iterating this process k times we will find that $U_kU_{k-1}\dots U_0$ is an $\epsilon(k)^2$ for approximation for U . To achieve the desired result, it suffices to stop the iteration when we find that $\epsilon(k)^2 < \epsilon$.

The final step is to count the number of elements of the approximation. Using that $U_k \in \mathcal{G}_{5^k\ell_0}$ we arrive to a total count of:

$$\sum_{n=0}^k 5^n \ell_0 = \ell_0 \frac{1 - 5^{k+1}}{1 - 5} = \ell_0 \frac{5^{k+1} - 1}{4} < \frac{5}{4} 5^k \ell_0 \quad (10)$$

From the expression of $\epsilon(k)$:

$$\left(\frac{3}{2}\right)^k < \frac{\log(\frac{1}{C^2\epsilon})}{2\log(\frac{1}{C\epsilon_0})} \quad (11)$$

Also it will be useful to consider this relation:

$$5^k = \left(\frac{3}{2}\right)^{kc} \Rightarrow c = \frac{\log 5}{\log 3/2} \approx 4 \quad (12)$$

Thus, if ℓ is the number of elements of the approximation:

$$\ell < \frac{5}{4} 5^k \ell_0 = \frac{5}{4} \left(\frac{3}{2}\right)^{kc} \ell_0 < \frac{5}{4} \left(\frac{\log(\frac{1}{C^2\epsilon})}{2\log(\frac{1}{C\epsilon_0})}\right)^c \ell_0 \quad (13)$$

Finally we achieve the desired result, \mathcal{G}_ℓ is an ϵ -net in $SU(d)$ for $\ell = \mathcal{O}(\log^c(1/\epsilon))$ with $c \approx 4$. \square

Remark 4.13. If we want to improve the result to $c > 3$ the Shrinking step will need to consider the quality of the nets and produce them in a different way. This procedure is called "Telescoping" and it will be commented later.

Most of the modifications of the previous theorem focus on lowering the constant c of the logarithm to the lower bound of 1. On the other hand, we can not forget the constants that appear in front of the logarithm like ϵ_0 or ℓ_0 . The constant ℓ_0 depends on the initial set of gates \mathcal{G} , and there is not a general construction to give an accurate value. Thus, another way of achieving efficient quantum circuits is studying the set of quantum gates. The constant ϵ_0 depends on the construction of the Shrinking Lemma and for an arbitrary dimension it is hard to calculate, but for the case of $SU(2)$ is it possible to find an upper bound.

4.4 The initial epsilon net

One of the most important conditions that the Solovay Kitaev theorem requires is the initial $(\epsilon_0, \epsilon_0^2)$ -net. So, it is important to give a numeric value for ϵ_0 because if one wishes to implement the theorem in a practical way, the value for ϵ_0 is fundamental.

In this section we are going to give an upper bound of ϵ_0 for the case of $SU(2)$. From the proof of the Solovay-Kitaev Theorem the value of ϵ_0 was given from the condition $C\epsilon_0 < 1$, thus we need to find the value of C . The constant C comes from the proof of the Shrinking lemma, specifically we left to calculate the constant c given by Proposition 4.4. In order to calculate the constant c we give the following proposition:

Proposition 4.14. *Let $\epsilon \geq 0$ and $y, z \in \mathbb{R}^3$ with $|y|, |z| \leq \epsilon$. Then*

$$\|[u(y), u(z)]_{gp} - u(y \times z)\| \leq c\epsilon^3,$$

$$\text{with } c \leq \frac{4}{3\sqrt{3}}.$$

Proof. Let $Y = \frac{1}{2}\sigma y$ and $Z = \frac{1}{2}\sigma z$, so $u(y) = e^{-iY}$ and $u(z) = e^{-iZ}$. By proposition 4.5 $e^{-[Y,Z]} = u(y \times z)$. Then:

$$\begin{aligned} [e^{-iY}, e^{-iZ}]_{gp} &= e^{-iY} e^{-iZ} e^{iY} e^{iZ} = \\ I - [Y, Z] + \frac{i}{2}[Y + Z, Y^2 + 2YZ + Z^2] + \mathcal{O}(\epsilon^4) \end{aligned}$$

and

$$e^{-[Y,Z]} = I - [Y, Z] + \mathcal{O}(\epsilon^4)$$

Putting all together it can be shown that:

$$T \equiv [e^{-iY} e^{iZ}]_{gp} - e^{-[Y,Z]} = -\frac{i}{4}[(y+z) \times (y \times z)] \cdot \sigma + \mathcal{O}(\epsilon^4)$$

The eigenvalues of T can be computed as $\lambda = \frac{1}{4}|(y+z) \times (y \times z)|$ (in fact it is a double eigenvalue). A more detailed computation can be found in [Har01] and [AR85]. Let θ be the angle between y and z :

$$\begin{aligned} |\lambda| &= \frac{1}{4}|(y+z)||y||z|\sin\theta| \\ &= \frac{1}{4}\sqrt{y^2 + z^2 + 2yz\cos\theta}yz|\sin\theta| \\ &\leq \frac{1}{4}\sqrt{2\epsilon^2(1 + \cos(\theta))\epsilon^2}|\sin(\theta)| = \frac{1}{4}\sqrt{2}\epsilon^3\sqrt{1 + \cos(\theta)}|\sin(\theta)| \\ &\leq \frac{2}{3\sqrt{3}}\epsilon^3 \end{aligned}$$

The last inequality comes from finding the maximum of the function $\sqrt{1 + \cos(\theta)}|\sin(\theta)|$ using standard techniques. In this case, the trace norm can be computed by simply summing all the eigenvalues of T :

$$\|[[u(y), u(z)]_{gp} - u(y \times z)]\| = \text{trace}\sqrt{TT^*} = |\lambda| + |\lambda| \leq \frac{4}{3\sqrt{3}}\epsilon^3$$

Thus $c \leq \frac{4}{3\sqrt{3}}$. □

From the expression of C that we mentioned in Section 4.2, $C = c+2$ we have $C \leq 2.77$. Since $C\epsilon_0 < 1$, the initial net will have a value for $\epsilon_0 < 1/C = 0.361$.

4.5 Generalization to SU(d)

Although the results given in the previous sections are stated for the case of SU(2), the Solovay-Kitaev Theorem can be generalized to the general case of SU(d), $d \geq 2$. The definitions of ϵ -net, (ϵ, ϵ') -net and S_ϵ trivially extend to arbitrary dimension. As in classical computation, we can generalize the concept of a gate too. A system of n qubits can be represented as a normalized vector in \mathbb{C}^{2n} , or more precisely as a point in \mathbb{CP}^{2^n-1} if we consider the equivalence between global phases. Thus, we can generalize the definition of a quantum gate acting on a single qubit to a quantum circuit acting on n -qubits as:

Definition 4.15. A *quantum gate* acting on n -qubits is a unitary transformation $U : \mathbb{C}\mathbb{P}^{n-1} \rightarrow \mathbb{C}\mathbb{P}^{n-1}$ such that $U \in U(2^n)$.

Note that the proof of the Solovay-Kitaev Theorem does not depend on the dimension, it depends uniquely on the Shrinking Lemma. Thus, if we generalize the Shrinking lemma the Solovay-Kitaev Theorem will follow. In order to generalize the Shrinking lemma, it suffices to prove only the first part of the proof given in Section 4.2, specifically:

Proposition 4.16. *Given a finite set of elements in $SU(d)$, $d \geq 2$, containing its own inverses such that $\langle G \rangle$ is dense in $SU(d)$ then:*

$$\text{If } \mathcal{G}_\ell \text{ is a } (\epsilon, \epsilon^2)\text{-net, then } \mathcal{G}_{4\ell} \text{ is a } (\epsilon^2, C\epsilon^3)$$

for some constant C .

The proof of this result relies on the Proposition 4.4 and the following proposition:

Proposition 4.17. *Let H be a $n \times n$ traceless Hermitian matrix. Then we can find F and G Hermitian matrices such that:*

$$[F, G] = iH \text{ with } \|F\|, \|G\|, \leq n^{\frac{1}{4}} \left(\frac{n-1}{2}\right)^{1/2} \sqrt{\|H\|}.$$

The proof of this Proposition, and the detailed construction of the general Shrinking Lemma can be found in Section 5 of [ND05]. In the next lines we are going to review the main idea in order to prove Proposition 4.16. Also we will need the generalization of Proposition 4.8 i.e given H an Hermitian matrix, then $d(I, \exp(iH)) = \|H\| + \mathcal{O}(\|H\|^3)$.

Proof. We sketch the main idea of the proof and refer to [ND05] for details. Given, any $U \in SU(d)$ we can find an hermitian matrix H such that $U = \exp(iH)$. If $U \in S_\epsilon^2$, using that $d(U, I) = \|H\| + \mathcal{O}(\|H\|^3)$ we can use Proposition 4.17 in order to find F and G such that $[F, G] = iH$ and $\|F\|, \|G\| \leq c'\epsilon$ where c' is the constant term given by the proposition. Setting $Y = \exp(iF)$ and $Z = \exp(iG)$ and using Proposition 4.4 it follows that

$$d(U, YZY^*Z^*) < C\epsilon^3$$

for some constant C . Moreover, $d(I, Y), d(I, Z) < c'\epsilon$ and since \mathcal{G}_ℓ is a (ϵ, ϵ^2) -net the result follows. The details for the constants can be found in the mentioned reference. \square

With that result, the second part of the Shrinking Lemma follows with the same construction. Thus, the Solovay-Kitaev Theorem can be generalized for an arbitrary dimension.

Theorem 4.18 (Solovay-Kitaev Theorem). *Let \mathcal{G} be a finite set of elements in $SU(d)$ with $d \geq 2$ containing its own inverses, such that $\langle \mathcal{G} \rangle$ is dense in $SU(d)$. Let $\epsilon > 0$ be a fixed integer. Then \mathcal{G}_ℓ is an ϵ -net in $SU(d)$ for*

$$\ell = \mathcal{O}(\log^c(1/\epsilon)) \text{ with } c = \frac{\log 5}{\log 3/2} \approx 4.$$

4.6 Controlling the nets

The general construction of the Solovay-Kitaev Theorem can be improved if we change some details of the original proof. In this section we are going to briefly review the modifications that Kitaev, Shen and Vyalıi add to the theorem in order to improve the number of gates to $\mathcal{O}(\log^c(1/\epsilon))$ with $c > 3$. The full construction can be found in [KN02]. In the next lines we are going to sketch some of the details. The main idea in order to improve the theorem is realizing that we have put "unnecessary" points in the nets and controlling a parameter called quality of the net. The following definitions state these concepts:

Definition 4.19. Let S be an ϵ -net for W in $SU(d)$. We say that S has no extra points if any $s \in S$ belongs to the ϵ -neighbourhood of W . The net S is called α -sparse ($0 < \alpha < 1$) if it has no extra points and the distance between any two distinct elements of S is greater than $\alpha\epsilon$.

Definition 4.20. Let S be an (ϵ, δ) -net. The quality of the net is defined as the ratio $q = \epsilon/\delta > 1$.

Since the quality plays an important role for the new constructions, we will denote a (ϵ, δ) -net with quality q as $(\epsilon, \epsilon/q)$ -net.

The first improvement for Theorem 4.18 modifies the step of Shrinking in order to work with sparse nets where there are not extra points. Let $S_1 \subseteq SU(n)$ and $S_2 \subseteq SU(n)$ be an $(\epsilon_1, \epsilon_1/q_1)$ -net and $(\epsilon_2, \epsilon_2/q_2)$ respectively for some quality $q_1, q_2 > 1$.

If $q_1 = q_2$ we will denote by $[[S_1, S_2]]_\alpha$ an α -sparse subnet selected from:

$$[[S_1, S_2]] = \{[s_1, s_2]_{gp} : s_1 \in S_1, s_2 \in S_2\}$$

Then the Shrinking lemma 4.2 is modified in the following way:

Lemma 4.21. If $q_1 = q_2 > 20$ and $\epsilon_1, \epsilon_2 \leq \mathcal{O}(1/q_1)$ then $[[S_1, S_2]]_{1/6}$ is an $\epsilon_1\epsilon_2/4, 5\epsilon_1\epsilon_2/q$ -net.

Note that this result is more precise than the original Shrinking Lemma 4.2. In fact, this result allows to combine different nets with the same quality and making a new one with the properties of being sparse. However, the quality of the new net is degraded by a factor of 20 from the original ones. The following proposition will solve this problem allowing to combine two nets into one of higher quality:

Proposition 4.22 (Telescoping). If $\epsilon_1/q_1 < \epsilon_2$, then the set

$$S_1 S_2 = \{s_1 s_2 : s_1 \in S_1, s_2 \in S_2\}$$

is an $(\epsilon_1, \epsilon_2/q_2)$ -net.

Combining two nets with this proposition is called telescoping the two nets. This procedure allows to combine the radius of one net with the error of the other.

The construction of the Solovay-Kitaev Theorem using the new Shrinking lemma differs from the one that we presented in the step of Shrinking. The new step is done by applying successively the new Shrinking Lemma to the initial net and then "telescoping" it with one of the previous nets in order to keep the original quality of the net. Since we are working with sparse nets the number of generators will be reduced and we will get a better result for the constant c . The exact construction of the new step can be found in the reference [\[KN02\]](#).

5 Conclusions

In this work, we have reviewed the standard proof of the Solovay-Kitaev Theorem based on the successive approximations of the ϵ -nets. There is an alternative proof in Chapter 13 of [KN02] but it requires a little more explanation of quantum gates and quantum circuits. In this work, we focused on presenting the Solovay-Kitaev Theorem for readers without prior knowledge on the field of quantum physics and quantum computing. We have tried to present the theorem in a self-contained way, since many works skip some of the details or leave them as exercises.

After the first proof of the Solovay-Kitaev Theorem, there have been several works extending the theorem. In [HBC02] it was proven that for a particular choice of the gate set, the optimal value of $c = 1$ can be achieved. Later, in [PS17], Sarnak and Parzanchevski discussed what they called Super Golden Gates, which are generators that give very efficient single quantum gates and can reduce the constant c to 1. These two works attack the problem of achieving efficient unitary approximations focusing on the gate set, instead of giving a general construction as the Solovay-Kitaev Theorem offers.

We would like to mention some other works related to the Solovay-Kitaev Theorem, which give other improvements, rather than achieving a better value of c . The first is an article by Freedman, Kitaev, and Lurie [FKL03]. In this work, it is proven that for any subset \mathcal{G} of a semisimple Lie group, there is a constant d , independent of the group, such that if all points of the group are within a distance less than d from the points of \mathcal{G} , then \mathcal{G} must generate a dense subset. This work relaxes the condition of having a universal set of quantum gates, to having a set that can approximate any element with a certain precision d . The second is a paper by Ozols and Bouland [AB17] where the condition of having the inverses in the initial set of the theorem is changed. Instead, they prove that if the set contains an irreducible representation of any finite group G , then, the Solovay-Kitaev theorem can be proven with a constant $c = \log_2 |G| + 3.97$. This is the first step to achieve a theorem without the condition of having the inverses, and it is important in order to define quantum complexity classes; problems that are in a complexity class must be independent of the gate set, thus the Solovay-Kitaev Theorem gives a polylogarithmic relation between operations of different sets which is acceptable. The last, is a paper by Kuperberg [Kup15]. This work is focused on the approximation of Jones Polynomials, a type of polynomials that appears in knot theory. It contains a generalization of the theorem for connected Lie groups whose Lie algebra is perfect. The theorem is used to approximate Tutte Polynomials which are graphs of polynomials that play an important role in graph theory. Note that from this last work, the Solovay-Kitaev Theorem can be used in other fields rather than quantum computing.

Finally, we summarize which parts of the Solovay-Kitaev Theorem remain as open problems:

- Prove if it is possible to lower the constant c of the theorem to the best value of $c = 1$ for any gate set.

- Achieve a free-inverse version of the theorem.
- Extending the theorem to other Lie groups.

So, if someone is interested, there is a lot of work to do beyond the Solovay-Kitaev Theorem!

References

- [AB17] Bouland A. and Ozols B., *Trading inverses for an irrep in the Solovay-Kitaev theorem.*, Leibniz International Proceedings in Informatics **111** (2017).
- [AR85] Horn R. A. and Johnson C. R., *Matrix analysis*, Cambridge University Press, Cambridge, 1985.
- [BMP⁺99] P.O. Boykin, T. Mor, M. Pulver, V. Roychowdhury, and F. Vatan, *On universal and fault-tolerant quantum computing: a novel basis and a new constructive proof of universality for Shor's basis*, Foundations of Computer Science, 1975., 16th Annual Symposium (1999), 486–494.
- [FH91] W. Fulton and J. Harris, *Representation theory*, Graduate Texts in Mathematics, vol. 129, Springer-Verlag, New York, 1991, A first course, Readings in Mathematics.
- [FKL03] M. H. Freedman, A. Kitaev, and J. Lurie, *Diameters of homogeneous spaces*, Math. Res. Lett. **10** (2003), no. 1, 11–20.
- [Hal15] B. Hall, *Lie groups, Lie algebras, and representations*, second ed., Graduate Texts in Mathematics, vol. 222, Springer, Cham, 2015.
- [Har01] A Harrow, *Quantum compiling. MIT undergraduated thesis.*
- [HBC02] A. W. Harrow, Recht B, and I. L. Chuang, *Efficient Discrete Approximations of Quantum Gates*, Journal of Mathematical Physics **43** (2002), 4445–4451.
- [Kit97] A. Y. Kitaev, *Quantum computations: algorithms and error correction*, Russian Mathematical Surveys **52** (1997), 1191–1249.
- [KN02] A. Yu. Shen A. H. Kitaev and Vyalyi M. N., *Classical and Quantum Computation*, Graduate Texts in Mathematics, vol. 47, American Mathematical Society, Providence, Rhode Island, 2002.
- [Kup15] G. Kuperberg, *How hard is to approximate the Jones polynomial?*, Theory of Computing **11** (2015), 183–219.
- [Lan99] S. Lang, *Fundamentals of differential geometry*, Graduate Texts in Mathematics, vol. 191, Springer-Verlag, New York, 1999.
- [Llo95] S. Lloyd, *Almost any quantum logic gate is universal*, Physical review letters **75** (1995), 346–349.
- [Mon16] A. Montanaro, *Quantum algorithms: an overview*, npj Quantum Information **2** (2016), 15023.

-
- [NC00] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge, 2000.
- [ND05] M. A. Nielsen and C. M. Dawson, *The Solovay-Kitaev algorithm*, *Quantum Information and Computation* **43** (2005), 81–95.
- [Ozo09] M. Ozols, *The Solovay-Kitaev theorem*, Preprint (2009).
- [Pre18] J. Preskill, *Quantum Computing in the NISQ era and beyond*, *Quantum* **2** (2018), 79.
- [PS17] O. Parzanchevski and P Sarnak, *Super-golden-gates for $PU(2)$* , *Advances in Mathematics* **327** (2017), 869–901.
- [Sar15] P Sarnak, *Letter to Aaronson and Pollington on the Solvay-Kitaev Theorem and golden gates*.
- [War83] F. W. Warner, *Foundations of differentiable manifolds and Lie groups*, Graduate Texts in Mathematics, vol. 94, Springer-Verlag, New York-Berlin, 1983, Corrected reprint of the 1971 edition.