The Solovay-Kitaev theorem

Maris Ozols

December 10, 2009

1 Introduction

There are several accounts of the Solovay-Kitaev theorem available [K97, NC00, KSV02, DN05]. I chose to base my report on [NC00], since it involves the Bloch sphere representation of the qubit which I like a lot. The drawback of my choice is that the proof of the Solovay-Kitaev theorem in [NC00] is somewhat sketchy and there are lots of gaps (mostly provided as exercises) left to be filled in. I tried to fill them in as much as I could and make my presentation self-contained and rigorous. However, there are still some details (indicated by gray notes on the margin of the page) that are either missing or I am not certain about.

1.1 Universal gate sets

Various sets of universal quantum gates are known. Among the most well known ones are:

- CNOT and all 1-qubit gates,
- CNOT, H, and T ($\frac{\pi}{8}$-gate),
- Toffoli and H (with ancillas can approximate any orthogonal matrix).

In the first case one can implement any quantum circuit exactly. However, from a practical point of view it may be more advantageous to consider a discrete universal gate set (i.e., one that can be used to approximate any quantum circuit), since then one can hope for a fault-tolerant implementation.

Since the number of gates required to perform a computation is used as a measure of complexity, it is a fundamental question whether one can translate a quantum circuit composed of gates from one universal set to another one without too much overhead.

1.2 Motivation

Assume that we are given a family of quantum circuits (e.g., for Grover's algorithm) consisting of CNOT gates and $f(n)$ 1-qubit gates, but we have to run it on a quantum computer, where CNOT is available, but instead of, say H and

T, another set $\mathcal{G}$ of 1-qubit is available. If we approximate each 1-qubit gate in the circuit with gates from $\mathcal{G}$ so that the error is at most $\varepsilon/f(n)$, then the overall error will be bounded by ε. We need something like $f(n)/\varepsilon$ gates to do this for each 1-qubit gate, so in total we need roughly $f(n)^2/\varepsilon$ gates.

However, the Solovay-Kitaev theorem states that we can get tolerance $\varepsilon/f(n)$ for 1-qubit gates only by using $O\big(\log^c(f(n)/\varepsilon)\big)$ gates from $\mathcal{G}$, so that the total cost of the algorithm is $O\big(f(n)\log^c(f(n)/\varepsilon)\big)$.

1.3 Definitions

1.3.1 Topology of metric spaces

Let X be a metric space and $d(\cdot,\cdot)$ be the corresponding metric.

Definition. Let $A, N \subset X$ (N finite) and $\varepsilon > 0$. N is called an *ε-net* for A if

$$\forall a \in A \;\; \exists p \in N : \; d(a,p) < \varepsilon. \tag{1}$$

Example. $\{0,1,2,3\}$ is a $(2/3)$-net for the interval $[0,3]$.

Definition. $D \subset X$ is said to be *dense* in X if

$$\forall x \in X \;\; \forall \varepsilon > 0 \;\; \exists p \in D : \; d(x,p) < \varepsilon. \tag{2}$$

Example. $\mathbb{Q}$ is dense in $\mathbb{R}$.

1.3.2 Trace norm

The only matrix norm that we will use throughout the paper is the *trace norm*. It is defined as follows:

$$\|A\| := \operatorname{Tr}|A| = \operatorname{Tr}\sqrt{A^\dagger A}. \tag{3}$$

Alternatively, $\|A\|$ is the sum of the singular values of A. If A is normal ($A^\dagger A = AA^\dagger$) and $\{\lambda_1,\ldots,\lambda_n\}$ are the eigenvalues of A, then $\|A\| = \sum_{i=1}^n |\lambda_i|$. The metric induced by the trace norm is given by $d(A,B) := \|A-B\|$.

Trace norm satisfies the following properties:

- unitary invariance: $\|UAV\| = \|A\|$ for any unitaries U and V,
- triangle inequality: $\|A+B\| \le \|A\| + \|B\|$,
- submultiplicativity: $\|AB\| \le \|A\|\,\|B\|$.

1.3.3 Free groups

Definition. The *free group generated by a set* $\mathcal{G}$ is the set $\langle\mathcal{G}\rangle$ of all finite sequences of symbols from $\big\{g, g^{-1} \mid g \in \mathcal{G}\big\}$ with concatenation of sequences (followed by cancellation) as the group operation.

Example. $\mathbb{Z}$ is free ($|\mathcal{G}| = 1$).

Notation. For any integer $l \geq 0$ let

$$\mathcal{G}^l := \{g_1^{\alpha_1} g_2^{\alpha_2} \dots g_l^{\alpha_l} \mid g_i \in \mathcal{G}, \alpha_i = \pm 1\}. \tag{4}$$

Note that $\langle \mathcal{G} \rangle = \mathcal{G}^0 \cup \mathcal{G}^1 \cup \mathcal{G}^2 \cup \dots$, where $\mathcal{G}^0 = \{\varepsilon\}$ and ε is the empty word.

Remark. If the generating set $\mathcal{G}$ is a subset of some group, then expression $g_1^{\alpha_1} g_2^{\alpha_2} \dots g_l^{\alpha_l}$ can naturally be interpreted in two ways. One can think of concatenation and inversion as formal operations on labels representing the elements of $\mathcal{G}$. However, one can use the underlying group structure instead and perform multiplication and inversion, respectively, as defined in the group. We will extensively make use of this twofold meaning.

2 "Shrinking" lemma

Let SU(2) stand for the *special unitary group* of 2×2 matrices, i.e., 2×2 unitary matrices of determinant 1. From now on we will assume that the generating set $\mathcal{G}$ is a finite subset of SU(2) that is closed under inverses and $\langle \mathcal{G} \rangle$ is dense in SU(2).

Notation. Let $S_\varepsilon := \{U \in \mathrm{SU}(2) : \|U - I\| < \varepsilon\}$ be an open ε-ball in SU(2) around the identity.

Lemma ("Shrinking" lemma)**.** There exist constants ε' and s, such that for any $\mathcal{G}$ and $\varepsilon \leq \varepsilon'$ we have:

$$\mathcal{G}^l \text{ is an } \varepsilon^2\text{-net for } S_\varepsilon \quad \Longrightarrow \quad \mathcal{G}^{5l} \text{ is an } s\varepsilon^3\text{-net for } S_{\sqrt{s\varepsilon^3}}.$$

The essence of this lemma is that it allows to construct better approximations at the cost of requiring longer sequences of generators and covering a smaller neighborhood of the identity. We defer its proof to Section 2.3.

To obtain approximations of desired accuracy, one has to apply this lemma several times. The following corollary just summarizes how the various parameters are affected by such iterative application.

Corollary (Iterated "shrinking" lemma)**.** There exist constants ε' and s, such that for any $\mathcal{G}$, $\varepsilon_0 \leq \varepsilon'$, and $k \in \mathbb{Z}_+$ we have:

$$\mathcal{G}^{l_0} \text{ is an } \varepsilon_0^2\text{-net for } S_{\varepsilon_0} \quad \Longrightarrow \quad \mathcal{G}^{l_k} \text{ is an } \varepsilon_k^2\text{-net for } S_{\varepsilon_k},$$

where $l_k := 5^k l_0$ and $\varepsilon_k := (s\varepsilon_0)^{(3/2)^k}/s$.

Proof. Let us repeatedly apply the "shrinking" lemma. One application allows us to transform the parameters according to the following map:

$$(l, \varepsilon^2, \varepsilon) \mapsto \left(5l, s\varepsilon^3, \sqrt{s\varepsilon^3}\right). \tag{5}$$

Clearly, after k iterations the first parameter transforms according to $l_0 \overset{k}{\mapsto} 5^k l_0$, as required. The second parameter is always the square of the third, so it remains to understand only the last parameter. We check that

$$\varepsilon_k \mapsto \sqrt{s\varepsilon_k^3} = \sqrt{s\left(\frac{(s\varepsilon_0)^{(3/2)^k}}{s}\right)^3} = \frac{(s\varepsilon_0)^{(3/2)^{k+1}}}{s} = \varepsilon_{k+1}, \tag{6}$$

as required. Thus, each application of the "shrinking" lemma increases k by 1. Hence, k applications gives us parameters $(l_k, \varepsilon_k^2, \varepsilon_k)$, as required. □

In the next two sections we will obtain some auxiliary results that will be used later (in Section 2.3) to prove the "shrinking" lemma.

2.1 Lie algebra of SU(2)

Let $\mathfrak{su}(n)$ denote the set of all $n \times n$ traceless Hermitian matrices. Here are two important facts about $\mathfrak{su}(n)$:

- If $H \in \mathfrak{su}(n)$, then $e^{-iH} \in \mathrm{SU}(n)$, since $\det e^{-iH} = e^{-i \operatorname{Tr} H} = 1$ and $e^{-iH}(e^{-iH})^\dagger = e^{-iH+iH} = I$.
- If $A, B \in \mathfrak{su}(n)$, then $i[A,B] \in \mathfrak{su}(n)$, since $(i[A,B])^\dagger = -i(BA - AB) = i[A,B]$ and $\operatorname{Tr}(i[A,B]) = i \operatorname{Tr}(AB - BA) = 0$.

Notation. Let $[A,B] := AB - BA$ and $[\![U,V]\!] := UVU^\dagger V^\dagger$ be the *additive (matrix)* and the *multiplicative (group) commutator*, respectively.

Claim 1. Let $A, B \in \mathfrak{su}(n)$ such that $\|A\|, \|B\| \le \varepsilon$ for some sufficiently small ε. Then there is a constant d such that $\left\|e^{-[A,B]} - [\![e^{-iA}, e^{-iB}]\!]\right\| \le d\varepsilon^3$.

Proof. The Taylor expansion for the first term is

$$e^{-[A,B]} = I - [A,B] + \frac{1}{2}[A,B]^2 - \dots. \tag{7}$$

To expand $[\![e^{-iA}, e^{-iB}]\!]$, note that

$$e^{-iA} = I - iA - \frac{A^2}{2} + \dots, \tag{8}$$

$$e^{+iA} = I + iA - \frac{A^2}{2} - \dots, \tag{9}$$

which gives

$$\begin{aligned} e^{-iA}e^{-iB}e^{iA}e^{iB} &= \left(I - iA - \frac{A^2}{2} + \dots\right)\left(I - iB - \frac{B^2}{2} + \dots\right) && (10)\\ &\quad \left(I + iA - \frac{A^2}{2} - \dots\right)\left(I + iB - \frac{B^2}{2} - \dots\right) && (11)\\ &= I + i^2(2AB - AB - BA - A^2 - B^2) - A^2 - B^2 + \dots && (12)\\ &= I - AB + BA + \dots && (13)\\ &= I - [A,B] - \dots. && (14) \end{aligned}$$

Note that expansions for both terms agree up to the second order. The claim follows by using the triangle inequality and submultiplicativity of the norm. □

Notation. We will use a formal inner product $\vec{r}\cdot\vec{\sigma}$ to denote the linear combination of Pauli matrices $\sigma_x := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $\sigma_y := \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$, and $\sigma_z := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ with coefficients given by the coordinates of $\vec{r} \in \mathbb{R}^3$, i.e., $\vec{r}\cdot\vec{\sigma} := r_x\sigma_x + r_y\sigma_y + r_z\sigma_z$.

Claim 2. Let $\vec{y}, \vec{z} \in \mathbb{R}^3$. Then $[\vec{y}\cdot\vec{\sigma}, \vec{z}\cdot\vec{\sigma}] = 2i(\vec{y}\times\vec{z})\cdot\vec{\sigma}$, where $\vec{y}\times\vec{z}$ is the cross product of vectors $\vec{y}$ and $\vec{z}$.

Proof. Consider the commutation relations of Pauli matrices: $[\sigma_j, \sigma_k]_l = 2i\varepsilon_{jkl}$, and the definition of the cross product on the standard basis: $(\vec{e}_j \times \vec{e}_k)_l = \varepsilon_{jkl}$, where ε_{jkl} is the totally antisymmetric Levi-Civita symbol. The claim follows by linearity. □

Notation. Let $u : \mathbb{R}^3 \to \mathrm{SU}(2)$ be the map $u(\vec{r}) := \exp(-\frac{i}{2}\vec{r}\cdot\vec{\sigma})$. It provides a correspondence between the Lie algebra $\mathfrak{su}(2)$ and the Lie group $\mathrm{SU}(2)$.

Claim 3. Let $\vec{y}, \vec{z} \in \mathbb{R}^3$. Then $\exp\big(-[\frac{1}{2}\vec{y}\cdot\vec{\sigma}, \frac{1}{2}\vec{z}\cdot\vec{\sigma}]\big) = u(\vec{y}\times\vec{z})$.

Proof. From Claim 2 we have: $\exp\big(-[\frac{1}{2}\vec{y}\cdot\vec{\sigma}, \frac{1}{2}\vec{z}\cdot\vec{\sigma}]\big) = \exp\big(-2i(\frac{1}{2}\vec{y}\times\frac{1}{2}\vec{z})\big) = \exp\big(-\frac{i}{2}(\vec{y}\times\vec{z})\big) = u(\vec{y}\times\vec{z})$. □

2.2 Distance relations

In this section we will relate the distance of elements from $\mathrm{SU}(2)$ to the distance of the corresponding vectors in $\mathbb{R}^3$.

Claim 4. Let $\vec{r} \in \mathbb{R}^3$. Then $\|u(\vec{r}) - I\| = 4\sin\frac{|\vec{r}|}{4}$.

Proof. Note that the eigenvalues of $\vec{r}\cdot\vec{\sigma}$ are $\pm|\vec{r}|$, because the characteristic polynomial of $\vec{r}\cdot\vec{\sigma}$ is

$$\begin{aligned}
\det(\vec{r}\cdot\vec{\sigma} - \lambda I) &= \begin{vmatrix} r_z - \lambda & r_x - ir_y \\ r_x + ir_y & -r_z - \lambda \end{vmatrix} && (15)\\
&= -(r_z - \lambda)(r_z + \lambda) - (r_x - ir_y)(r_x + ir_y) && (16)\\
&= -(r_z^2 - \lambda^2) - (r_x^2 - i^2 r_y^2) && (17)\\
&= \lambda^2 - r_x^2 - r_y^2 - r_z^2 = 0. && (18)
\end{aligned}$$

Thus, the eigenvalues of $u(\vec{r}) - I$ are $e^{\pm\frac{i}{2}|\vec{r}|} - 1$. A simple computation gives

$$\begin{aligned}
\left|e^{\pm\frac{i}{2}|\vec{r}|} - 1\right| &= \sqrt{\left(\cos\tfrac{|\vec{r}|}{2} - 1\right)^2 + \left(\sin\tfrac{|\vec{r}|}{2}\right)^2} && (19)\\
&= \sqrt{\left(\cos^2\tfrac{|\vec{r}|}{2} - 2\cos\tfrac{|\vec{r}|}{2} + 1\right) + \left(1 - \cos^2\tfrac{|\vec{r}|}{2}\right)} && (20)\\
&= \sqrt{2 - 2\cos\tfrac{|\vec{r}|}{2}} = 2\sqrt{\frac{1}{2}\left(1 - \cos\tfrac{|\vec{r}|}{2}\right)} = 2\sin\frac{|\vec{r}|}{4}, && (21)
\end{aligned}$$

hence $\|u(\vec{r}) - I\| = \left|e^{+\frac{i}{2}|\vec{r}|} - 1\right| + \left|e^{-\frac{i}{2}|\vec{r}|} - 1\right| = 4\sin\frac{|\vec{r}|}{4}$. □

Claim 5. Let $\vec{r} \in \mathbb{R}^3$. If $u(\vec{r}) \in S_\varepsilon$ then $|\vec{r}| < \varepsilon + O(\varepsilon^3)$.

Proof. From Claim 4 we have: $\|u(\vec{r}) - I\| = 4 \sin \frac{|\vec{r}|}{4} < \varepsilon$. Thus, $|\vec{r}| < 4 \arcsin \frac{\varepsilon}{4}$. Result follows from the Taylor expansion $\arcsin z = z + \frac{1}{2} \cdot \frac{z^3}{3} + \frac{1 \cdot 3}{2 \cdot 4} \cdot \frac{z^5}{5} + \ldots$. □

Claim 6. If $\vec{y}, \vec{z} \in \mathbb{R}^3$ and $|\vec{y}|, |\vec{z}| < \varepsilon$ then $\|u(\vec{y}) - u(\vec{z})\| = |\vec{y} - \vec{z}| + O(\varepsilon^3)$.

Proof. Using unitary invariance and triangle inequality we get:

$$\|u(\vec{y}) - u(\vec{z})\| = \|u(\vec{y})u(\vec{z})^\dagger - I\| \tag{22}$$

$$\leq \|u(\vec{y})u(\vec{z})^\dagger - u(\vec{y} - \vec{z})\| + \|u(\vec{y} - \vec{z}) - I\|. \tag{23}$$

We have to make sure that the first term is small enough. Then applying Claim 4 to the second term we get $\|u(\vec{y} - \vec{z}) - I\| = 4 \sin \frac{|\vec{y} - \vec{z}|}{4}$ and result follows from the Taylor expansion $\sin \alpha = \alpha - \frac{\alpha^3}{3!} + \frac{\alpha^5}{5!} - \ldots$. □

How to bound the 1st term?

2.3 Proof of the "shrinking" lemma

Recall the statement of the "shrinking" lemma:

Lemma ("Shrinking" lemma)**.** There exist constants ε' and s, such that for any $\mathcal{G}$ and $\varepsilon \leq \varepsilon'$ we have:

$$\mathcal{G}^l \text{ is an } \varepsilon^2\text{-net for } S_\varepsilon \quad \Longrightarrow \quad \mathcal{G}^{5l} \text{ is an } s\varepsilon^3\text{-net for } S_{\sqrt{s\varepsilon^3}}.$$

Proof. To prove this lemma, we have to transform the parameters of the net according to the map $(l, \varepsilon^2, \varepsilon) \mapsto \left(5l, s\varepsilon^3, \sqrt{s\varepsilon^3}\right)$. Let us first show how to obtain parameters $\left(4l, s\varepsilon^3, \varepsilon^2\right)$.

Given $U \in S_{\varepsilon^2}$ pick $\vec{x} \in \mathbb{R}^3$ such that

$$U = u(\vec{x}), \quad |\vec{x}| < \varepsilon^2 + O(\varepsilon^6), \tag{24}$$

where the inequality comes from Claim 5. Next, choose $\vec{y}, \vec{z} \in \mathbb{R}^3$ such that

$$\vec{x} = \vec{y} \times \vec{z}, \quad |\vec{y}|, |\vec{z}| < \varepsilon + O(\varepsilon^5). \tag{25}$$

(This is possible because $|\vec{x}| = |\vec{y} \times \vec{z}| \leq |\vec{y}|\,|\vec{z}|$.) Let us approximate $u(\vec{y})$ and $u(\vec{z})$ by elements of $\mathcal{G}^l$. We use Claim 5 to make sure that $u(\vec{y}), u(\vec{z}) \in S_\varepsilon$. Since $\mathcal{G}^l$ is an ε^2-net for S_ε, we can choose $\vec{y}_0, \vec{z}_0 \in \mathbb{R}^3$ such that $u(\vec{y}_0), u(\vec{z}_0) \in \mathcal{G}^l \cap S_\varepsilon$ and they are good approximations of $u(\vec{y})$ and $u(\vec{z})$, respectively:

$$\|u(\vec{y}_0) - I\| < \varepsilon, \quad \|u(\vec{y}_0) - u(\vec{y})\| < \varepsilon^2 + O(\varepsilon^5), \tag{26}$$

$$\|u(\vec{z}_0) - I\| < \varepsilon, \quad \|u(\vec{z}_0) - u(\vec{z})\| < \varepsilon^2 + O(\varepsilon^5). \tag{27}$$

I don't think we need the $O(\varepsilon^5)$ term.

We apply Claims 5 and 6 to the first and second column, respectively:

$$|\vec{y}_0| < \varepsilon + O(\varepsilon^3), \quad |\vec{y}_0 - \vec{y}| < \varepsilon^2 + O(\varepsilon^3), \tag{28}$$

$$|\vec{z}_0| < \varepsilon + O(\varepsilon^3), \quad |\vec{z}_0 - \vec{z}| < \varepsilon^2 + O(\varepsilon^3). \tag{29}$$

$O(\varepsilon^3)$ in the 2nd column is just a guess.

We want to show that $\|u(\vec{x}) - [\![u(\vec{y}_0), u(\vec{z}_0)]\!]\| < s\varepsilon^3$ for some constant s, i.e., U can be approximated well enough using a sequence of $4l$ elements from $\mathcal{G}$ (they are needed to implement the group commutator $[\![u(\vec{y}_0), u(\vec{z}_0)]\!]$). Let us use the triangle inequality

$$\|u(\vec{x}) - [\![u(\vec{y}_0), u(\vec{z}_0)]\!]\| \tag{30}$$

$$\leq \underbrace{\|u(\vec{x}) - u(\vec{y}_0 \times \vec{z}_0)\|}_{D_1} + \underbrace{\|u(\vec{y}_0 \times \vec{z}_0) - [\![u(\vec{y}_0), u(\vec{z}_0)]\!]\|}_{D_2} \tag{31}$$

and consider both terms separately.

For the first term we use Claim 6 (note that $|\vec{y} \times \vec{z}| < \varepsilon^2 + O(\varepsilon^6)$ and $|\vec{y}_0 \times \vec{z}_0| < \varepsilon^2 + O(\varepsilon^4)$):

$$D_1 = \|u(\vec{y} \times \vec{z}) - u(\vec{y}_0 \times \vec{z}_0)\| \tag{32}$$

$$= |\vec{y} \times \vec{z} - \vec{y}_0 \times \vec{z}_0| + O(\varepsilon^6) \tag{33}$$

$$= \big|\big((\vec{y} - \vec{y}_0) + \vec{y}_0\big) \times \big((\vec{z} - \vec{z}_0) + \vec{z}_0\big) - \vec{y}_0 \times \vec{z}_0\big| + O(\varepsilon^6) \tag{34}$$

$$= |(\vec{y} - \vec{y}_0) \times (\vec{z} - \vec{z}_0) + \vec{y}_0 \times (\vec{z} - \vec{z}_0) + (\vec{y} - \vec{y}_0) \times \vec{z}_0| + O(\varepsilon^6) \tag{35}$$

$$\leq |(\vec{y} - \vec{y}_0) \times (\vec{z} - \vec{z}_0)| + |\vec{y}_0 \times (\vec{z} - \vec{z}_0)| + |(\vec{y} - \vec{y}_0) \times \vec{z}_0| + O(\varepsilon^6) \tag{36}$$

$$\leq |\vec{y} - \vec{y}_0|\,|\vec{z} - \vec{z}_0| + |\vec{y}_0|\,|\vec{z} - \vec{z}_0| + |\vec{y} - \vec{y}_0|\,|\vec{z}_0| + O(\varepsilon^6) \tag{37}$$

$$< c\varepsilon^3 + O(\varepsilon^4) \tag{38}$$

for some constant c. For the second term we use Claim 3 and then Claim 1 with a slightly larger constant d', since $|\vec{y}_0|$ and $|\vec{z}_0|$ can be larger than ε:

$$D_2 = \big\|\exp\big(-[\tfrac{1}{2}\vec{y}_0 \cdot \vec{\sigma}, \tfrac{1}{2}\vec{z}_0 \cdot \vec{\sigma}]\big) - [\![\exp(-\tfrac{i}{2}\vec{y}_0 \cdot \vec{\sigma}), \exp(-\tfrac{i}{2}\vec{z}_0 \cdot \vec{\sigma})]\!]\big\| \leq d'\varepsilon^3. \tag{39}$$

When we put both terms together, we get $\|U - [\![u(\vec{y}_0), u(\vec{z}_0)]\!]\| \leq s\varepsilon^3$ for some constant s.

It remains to increase the size of the neighborhood that is covered by using one more sequence of l generators, i.e., we want to transform the parameters $\big(4l, s\varepsilon^3, \varepsilon^2\big)$ to $\big(5l, s\varepsilon^3, \sqrt{s\varepsilon^3}\big)$. Recall that the initial parameters were $(l, \varepsilon^2, \varepsilon)$ meaning that $\mathcal{G}^l$ is an ε^2-net for S_ε. Thus, for given $U \in S_{\sqrt{s\varepsilon^3}}$ we can find $V \in \mathcal{G}^l$ such that $\|U - V\| = \big\|UV^\dagger - I\big\| < \varepsilon^2$ meaning that $UV^\dagger \in S_{\varepsilon^2}$. Since $\mathcal{G}^{4l}$ is an $s\varepsilon^3$-net for S_{ε^2}, we can use the first part to find $\vec{y}_0, \vec{z}_0 \in \mathbb{R}^3$ such that $\big\|UV^\dagger - [\![u(\vec{y}_0), u(\vec{z}_0)]\!]\big\| \leq s\varepsilon^3$, which completes the proof. □

Why does $\sqrt{s\varepsilon^3} \leq \varepsilon$?

3 Solovay-Kitaev theorem

Recall that $\mathcal{G}$ is a finite subset of SU(2) that is closed under inverses and $\langle\mathcal{G}\rangle$ is dense in SU(2).

Theorem (Solovay-Kitaev theorem)**.** There is a constant c such that for any $\mathcal{G}$ and $\varepsilon > 0$ one can choose $l = O\big(\log^c(1/\varepsilon)\big)$ so that $\mathcal{G}^l$ is an ε-net for SU(2).

The iterated "shrinking" lemma allows to obtain a good approximation for any element of SU(2) that is sufficiently close to identity. To prove the Solovay-Kitaev theorem, we have to obtain a good approximation for *any* element of SU(2). This is done by starting with a rough approximation and then refining it by invoking the iterated "shrinking" lemma for different values of k (starting with smaller ones). Intuitively this corresponds to approaching the desired element by performing steps whose size decreases the closer we get. We adapt the step size, because larger steps can be implemented using shorter sequences of generators. In [KSV02] this procedure is referred to as "zooming in".

Proof (of the **Solovay-Kitaev theorem**). Let us choose the initial value ε_0 so that $\varepsilon_0 < \varepsilon'$ to be able to apply the iterated "shrinking" lemma. In addition we want $s\varepsilon_0 < 1$ to make sure that ε_k decreases as we increase k. Moreover, we also make sure that ε_0 is small enough so that $\varepsilon_k^2 < \varepsilon_{k+1}$.

Since $\langle \mathcal{G} \rangle$ is dense in SU(2), we can choose l_0 large enough[1] so that $\mathcal{G}^{l_0}$ is an ε_0^2-net for SU(2) (and hence for S_{ε_0} as well) no matter how small ε_0 is. Thus, given any $U \in \mathrm{SU}(2)$, we can choose $U_0 \in \mathcal{G}^{l_0}$ such that $\|U - U_0\| < \varepsilon_0^2$. Let $\Delta_1 := UU_0^\dagger$ be the "difference" of U and U_0. Then

$$\|\Delta_1 - I\| = \|(U - U_0)U_0^\dagger\| = \|U - U_0\| < \varepsilon_0^2 < \varepsilon_1. \tag{40}$$

Hence, $\Delta_1 \in S_{\varepsilon_1}$. By invoking the iterated "shrinking" lemma with $k = 1$, there exists $U_1 \in \mathcal{G}^{l_1}$ such that $\|\Delta_1 - U_1\| = \|UU_0^\dagger - U_1\| = \|U - U_1U_0\| < \varepsilon_1^2$.

Similarly, let $\Delta_2 := \Delta_1 U_1^\dagger = UU_0^\dagger U_1^\dagger$. Then

$$\|\Delta_2 - I\| = \|(U - U_1U_0)U_0^\dagger U_1^\dagger\| = \|U - U_1U_0\| < \varepsilon_1^2 < \varepsilon_2. \tag{41}$$

Thus, $\Delta_2 \in S_{\varepsilon_2}$ and we can invoke the iterated "shrinking" lemma (with $k = 2$ this time) to get $U_2 \in \mathcal{G}^{l_2}$ such that $\|\Delta_2 - U_2\| = \|UU_0^\dagger U_1^\dagger - U_2\| = \|U - U_2U_1U_0\| < \varepsilon_2^2$.

If we continue in this way, after k steps we get $U_k \in \mathcal{G}^{l_k}$ such that

$$\|U - U_kU_{k-1} \cdots U_0\| < \varepsilon_k^2. \tag{42}$$

Thus, we have obtained a sequence of

$$L = \sum_{m=0}^{k} l_m = \sum_{m=0}^{k} 5^m l_0 = \frac{5^{k+1} - 1}{4} l_0 < \frac{5}{4} 5^k l_0 \tag{43}$$

gates that approximates U to accuracy ε_k^2. To determine the value of k, we set $\varepsilon_k^2 = \left((s\varepsilon_0)^{(3/2)^k}/s\right)^2 = \varepsilon$ and solve for k:

$$\left(\frac{3}{2}\right)^k = \frac{\log(1/s^2\varepsilon)}{2\log(1/s\varepsilon_0)}. \tag{44}$$

[1] Note that $\langle \mathcal{G} \rangle$ is an ε_0^2-net for SU(2) as it is dense in SU(2). Thus, the ε_0^2-neighborhoods of all elements from $\langle \mathcal{G} \rangle$ form an open cover $\mathcal{C}$ of SU(2). Since SU(2) is compact, $\mathcal{C}$ has a finite subcover $\mathcal{C}'$. Clearly, $\mathcal{C}' \subseteq \mathcal{G}^{l_0}$ for some finite value of l_0 as $\mathcal{C}'$ is finite.

Note that we can always choose ε_0 slightly smaller so that the obtained value of k is an integer.[2] Let $c = \log 5/\log(3/2) \approx 3.97$ so that $5^k = (\frac{3}{2})^{kc}$. Then

$$L < \frac{5}{4}5^k l_0 = \frac{5}{4}\left(\frac{3}{2}\right)^{kc} l_0 = \frac{5}{4}\left(\frac{\log(1/s^2\varepsilon)}{2\log(1/s\varepsilon_0)}\right)^c l_0. \tag{45}$$

Hence, for any $U \in \mathrm{SU}(2)$ there is a sequence of $L = O\big(\log^c(1/\varepsilon)\big)$ gates that approximates U to accuracy ε. □

References

[K97] Kitaev A.Yu., Quantum computations: algorithms and error correction, UMN, **52**:6(318), pp. 53–112 (1997), http://mi.mathnet.ru/eng/umn/v52/i6/p53.

[NC00] Nielsen M.A., Chuang I.L., Quantum Computation and Quantum Information (Cambridge University Press, 2000), Appendix 3, pp. 617–624.

[KSV02] Kitaev A.Y., Shen A.H., Vyalyi M.N., Classical and quantum computation (AMS, 2002), §8.3., pp. 75–82.

[DN05] Dawson C.M., Nielsen M.A., The Solovay-Kitaev algorithm, arXiv:quant-ph/0505030v2.

[2] We want to have equality in equation (44) instead of inequality, since the sign in equation (A3.3) in [NC00] is wrong.