UNIVERSIDAD
DE GRANADA

Facultad de Ciencias
Escuela Técnica Superior de Ingeniería Informática y Telecomunicaciones

Grado en Ingeniería Informática y Matemáticas

TRABAJO DE FIN DE GRADO

Aprendizaje Automático Cuántico para predicción de series temporales

Presentado por:
Nicolás Segura Kunsági

Tutores:
Manuel PegalajarCuéllar
Ciencias de la Computación e Inteligencia Artificial

Gabriel Navarro Garulo
Ciencias de la Computación e Inteligencia Artificial

Curso académico 2022-2023

Aprendizaje Automático Cuántico para predicción de series temporales

Nicolás Segura Kunsági

Nicolás Segura Kunsági *Aprendizaje Automático Cuántico para predicción de series temporales*. Trabajo de fin de Grado. Curso académico 2022-2023.

Responsable de tutorización

Manuel PegalajarCuéllar
Ciencias de la Computación e Inteligencia Artificial

Gabriel Navarro Garulo
Ciencias de la Computación e Inteligencia Artificial

Declaración de originalidad

D. Nicolás Segura Kunsági

Declaro explícitamente que el trabajo presentado como Trabajo de Fin de Grado (TFG), correspondiente al curso académico 2022-2023, es original, entendida esta, en el sentido de que no ha utilizado para la elaboración del trabajo fuentes sin citarlas debidamente.

En Granada a 25 de mayo de 2023

Fdo: Nicolás Segura Kunsági

Índice general

Summary

An english summary of the project (around 800 and 1500 words are recommended). Hasta no tener las partes del TFG no haré un resumen del trabajo.

Introducción

Lo que hay que hacer: De acuerdo con la comisión de grado, el TFG debe incluir una introducción en la que se describan claramente los objetivos previstos inicialmente en la propuesta de TFG, indicando si han sido o no alcanzados, los antecedentes importantes para el desarrollo, los resultados obtenidos, en su caso y las principales fuentes consultadas.

Nicolás: Fuentes consultadas [McM07], [Hid19] y [Sut19]

Para aprendizaje automatico [SP18], [ZJQ23]

Para los diagramas se ha usado el paquete quantikz y qiskit, documentado cada uno en [Kay19] y [Qis23]

Parte I.

Fundamentos de la Computación Cuántica y Qiskit

Daremos las definiciones y operaciones básicas para la computación cuántica. En paralelo a la teoría mostraremos ejemplos en Qiskit de una posible implementación de dichos conceptos.

1. Qubit

En el presente capítulo se definirá el concepto de qubit, algunas propiedades suyas y una representación gráfica. Luego veremos como trabajamos con varios qubits juntos y la propiedad de entrelazamiento.

1.1. ¿Qué es un qubit?

En computación clásica la unidad básica de información es el bit y este solo puede tomar como valores 0 y 1. Análogamente en computación cuántica la unidad básica es el quantum bit (qubit) y además de tomar como valores el "0" y el "1" puede tomar valores que sean combinación entre el "0" y el "1". Concretemos formalmente esta idea.

Definición 1.1 (Qubit)**.** Un qubit es un vector de norma 1 en el espacio de Hilbert $\mathbb{C}^2$ sobre el cuerpo de los complejos. Denotaremos al conjunto de todos los qubits como $\mathcal{Q}$.

La norma que usamos es la inducida por el siguiente producto escalar, supongamos $u = \begin{pmatrix} a \\ b \end{pmatrix}, v = \begin{pmatrix} c \\ d \end{pmatrix} \in \mathbb{C}^2$, entonces

$$< u|v >=< \begin{pmatrix} a \\ b \end{pmatrix} | \begin{pmatrix} c \\ d \end{pmatrix} >= ac + bd.$$

Repasemos las propiedades del producto escalar en espacios complejos:

1. $< u|u > \geq 0$ y $< v|v > = 0 \Leftrightarrow v = 0, \forall u, v \in \mathbb{C}^2$.
2. $< u|v > = \overline{< v|u >}, \forall u, v \in \mathbb{C}^2$.
3. $< u|av + bw > = a < u|v > +b < u|w >, \forall u, v, w \in \mathbb{C}^2, \forall a, b \in \mathbb{C}$, linealidad en la segunda componente.
4. $< au + bv|w > = \overline{a} < u|w > +\overline{b} < v|w >, \forall u, v, w \in \mathbb{C}^2, \forall a, b \in \mathbb{C}$, antilinealidad en la primera componente.

La norma inducida por un producto escalar se define como: $\|u\| = \sqrt{< u|u >}$.

Proposición 1.1. *Sea $B = \{v, w\}$ una base ortonormal de $\mathbb{C}^2$. Entonces u es un vector de norma 1 si y solo si podemos escribir $u = av + bw$ donde $|a|^2 + |b|^2 = 1$.*

Demostración. Como B es una base de $\mathbb{C}^2$ entonces existen unos únicos $a, b \in \mathbb{C}$ tales que $u = av + bw$. Solo nos bastaría comprobar $\|u\| = 1 \Leftrightarrow |a|^2 + |b|^2 = 1$.
Y por definición $\|u\| = \sqrt{< u|u >}$, podemos expresar:

$$\begin{aligned} < u|u > &=< av + bw|av + bw >= a < av + bw|v > +b < av + bw|w > \\ &= a\overline{a} < v|v > +a\overline{b} < w|v > +b\overline{a} < v|w > +b\overline{b} < w|w >= a\overline{a} + b\overline{b} \\ &= |a|^2 + |b|^2. \end{aligned}$$

.

Teniendo en cuenta la igualdad anterior, entonces $1 = \|u\| = \sqrt{< u|u >} \Leftrightarrow 1 =< u|u >= |a|^2 + |b|^2$. □

Tomemos la base usual de $\mathbb{C}^2$, $B_u = \{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\}$. Esta base cumple las propiedades de la proposición anterior, además vamos a denotar sugestivamente a los vectores de la base como

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

entonces, dado un qubit, $|\psi\rangle$, podemos ponerlo en combinación lineal de la base

$$|\psi\rangle = a\,|0\rangle + b\,|1\rangle\,.$$

Donde $|a|^2 + |b|^2 = 1$. Ejemplos concretos de qubits son:

$$|0\rangle \ , \ |1\rangle \ , \ \frac{1}{\sqrt{2}}\,|0\rangle + \frac{-i}{\sqrt{2}}\,|1\rangle \ , \ \frac{1+i}{2}\,|0\rangle + \frac{1-i}{2}\,|1\rangle\,.$$

Demos un significado a a y a b. Un qubit se encuentra en superposición, una combinación lineal entre $|0\rangle$ y $|1\rangle$, cuando tratamos de medir su valor inevitablemente colapsa, es decir se proyecta sobre $|0\rangle$ o $|1\rangle$. Este colapso se representa como proyección donde luego normalizamos el resultado para que siga teniendo módulo 1. Una vez colapsado, $|0\rangle$ y $|1\rangle$ podemos representarlos como 0 y 1 usando bits clásicos.

Si interpretamos la medición de un qubit físicamente, como medimos un qubit la probabilidad de medirlo es 1,

$$\begin{aligned}1 = P(|\psi\rangle) &= \||\psi\rangle\|^2 =< a\,|0\rangle + b\,|1\rangle\,, a\,|0\rangle + b\,|1\rangle > \\ &= |a|^2\||0\rangle\|^2 + |b|^2\||1\rangle\|^2 + (\bar{a}b < |0\rangle\,, |1\rangle > + a\bar{b} < |1\rangle\,, |0\rangle >) \\ &= |a|^2\||0\rangle\|^2 + |b|^2\||1\rangle\|^2 = |a|^2 P(|0\rangle) + |b|^2 P(|1\rangle) = |a|^2 + |b|^2.\end{aligned}$$

La medición de un qubit implica que colapsara en uno de los estados que lo conforman. Estos estados son $|0\rangle$ y $|1\rangle$, y la probabilidad de llegar a cada estado es la probabilidad de medir dicho estado por si solo por el coeficiente que le acompaña. Tal y como lo tenemos descrito, dichas probabilidades son $|a|^2$ y $|b|^2$ para $|0\rangle$ y $|1\rangle$ respectivamente.

Como observación, al descomponer un qubit obtuvimos que un término que se anuló. Si lo interpretamos en otra base, este término influye a la hora de una medición. A esto se le conoce como interferencia y representa la posibilidad de llegar a estados desconocidos. Por este motivo veremos que las mediciones se realizan respecto a bases ortonormales con el fin de eliminar incertidumbre.

1.1.1. Notación braket

En computación cuántica en vez de escribir los qubits como vectores usamos la notación en brakets. En esta sección vamos a detallar está notación y unos ejemplos para familiarizarnos con ella, ya que a partir de ahora se usará con más frecuencia.

Supongamos que el vector en la base usual $\begin{pmatrix} a \\ b \end{pmatrix}$, entonces lo denotamos como un ket

$$|\psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix} = a\,|0\rangle + b\,|1\rangle\,.$$

Para la definición de un bra nos basaremos en el siguiente teorema.

Teorema 1.1 (Teorema de representación de Riesz)**.** *Sea H un espacio de Hilbert, H' su espacio dual. Entonces para todo $f \in H'$, existe un único $v_f \in H$ tal que:*

$$f(x) =< v_f, x >\,,\ \forall x \in H.$$

Aplicando este teorema al espacio $\mathbb{C}^2$, sea una aplicación lineal, $f : \mathbb{C}^2 \to \mathbb{C}$, entonces existe un único $x \in \mathbb{C}^2$, tal que $f(y) =< x, y >$. Si dicho x tiene norma uno entonces puedo considerar que es un qubit $|\psi\rangle = x$. Finalmente definimos un bra como:

$$\langle\psi| = f.$$

A nivel de cálculo práctico, el bra de un qubit $|\psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix}$ es

$$\langle\psi| = \overline{|\psi\rangle^T} = \begin{pmatrix} \overline{a} & \overline{b} \end{pmatrix} = \overline{a}\,\langle 0| + \overline{b}\,\langle 1|\,.$$

Basándonos en la definición de bra podemos escribir el producto escalar entre dos qubits $|\psi\rangle = \begin{pmatrix} a \\ b \end{pmatrix}$, $|\phi\rangle = \begin{pmatrix} c \\ d \end{pmatrix}$ como

$$< \begin{pmatrix} a \\ b \end{pmatrix} | \begin{pmatrix} c \\ d \end{pmatrix} >= \langle\psi|\,|\phi\rangle = \langle\psi|\phi\rangle\,,$$

y si invertimos el orden del ket y del bra

$$|\psi\rangle\,\langle\phi| = \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} \overline{c} & \overline{d} \end{pmatrix} = \begin{pmatrix} a\overline{c} & a\overline{d} \\ b\overline{c} & b\overline{d} \end{pmatrix}.$$

Dejando todo en base a los qubits $|0\rangle$,$|1\rangle$ obtenemos

$$\begin{aligned} \langle\psi|\phi\rangle &= (\overline{a}\,\langle 0| + \overline{b}\,\langle 1|)(c\,|0\rangle + d\,|1\rangle) = \overline{a}c\,\langle 0|\,|0\rangle + \overline{a}d\,\langle 0|\,|1\rangle + \overline{b}c\,\langle 1|\,|0\rangle + \overline{b}d\,\langle 1|\,|1\rangle \\ |\psi\rangle\,\langle\phi| &= (a\,|0\rangle + b\,|1\rangle)(\overline{c}\,\langle 0| + \overline{d}\,\langle 1|) = a\overline{c}\,|0\rangle\,\langle 0| + a\overline{d}\,|0\rangle\,\langle 1| + b\overline{c}\,|1\rangle\,\langle 0| + b\overline{d}\,|1\rangle\,\langle 1|\,. \end{aligned}$$

1.2. Esfera de Bloch

Hemos visto que podemos escribir un qubit, o el estado cuántico de un qubit, de la forma $|\psi\rangle = a\,|0\rangle + b\,|1\rangle$ donde a, b indican la probabilidad de medir $|0\rangle\,, |1\rangle$ respectivamente. Si multiplicamos por $e^{i\phi}$ a nuestro qubit $|\psi\rangle$, $e^{i\phi}\,|\psi\rangle = e^{i\phi}a\,|0\rangle + e^{i\phi}b\,|1\rangle$ cambiamos el estado cuántico pero mantemos las probabilidades de medir $|0\rangle\,, |1\rangle$ ya que $|e^{i\phi}a|^2 = |a|^2$, $|e^{i\phi}b|^2 = |b|^2$.

Con la observación anterior y la necesidad de medir un qubit para poder usarlo, concluimos que aunque dos qubits diferentes representen dos estados distintos, estos estados son el mismo a nivel práctico si las probabilidades de medición para $|0\rangle$ y $|1\rangle$ coinciden para ambos

estados. Por ejemplo, sean los estados

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{-i}{\sqrt{2}}|1\rangle \ , \ |\psi_2\rangle = \frac{1+i}{2}|0\rangle + \frac{1-i}{2}|1\rangle .$$

Podemos decir que son el mismo porque $|\frac{1}{\sqrt{2}}|^2 = \frac{1}{2} = |\frac{1+i}{2}|^2$ y $|\frac{-i}{\sqrt{2}}|^2 = \frac{1}{2} = |\frac{1-i}{2}|^2$.

Basándonos este concepto es razonable definir la siguiente relación, R, en $\mathcal{Q}$.

$$|\psi_1\rangle \ R \ |\psi_2\rangle \Leftrightarrow \exists e^{i\phi} : e^{i\phi}|\psi_1\rangle = |\psi_2\rangle .$$

Proposición 1.2. *La relación R es una relación de equivalencia.*

Demostración. Veamos que R es transitiva, reflexiva y simétrica.

- R es transitiva: supongamos $|\psi_1\rangle \ R \ |\psi_2\rangle$,$|\psi_2\rangle \ R \ |\psi_3\rangle$, entonces existen ϕ_1,ϕ_2 tales que $e^{i\phi_1}|\psi_1\rangle = |\psi_2\rangle$,$e^{i\phi_2}|\psi_2\rangle = |\psi_3\rangle$.

  Definimos $\phi_3 = \phi_2 + \phi_1$, luego:

  $$e^{i\phi_3}|\psi_1\rangle = e^{i\phi_2 + i\phi_1}|\psi_1\rangle = e^{i\phi_2}e^{i\phi_1}|\psi_1\rangle = e^{i\phi_2}|\psi_2\rangle = |\psi_3\rangle .$$

  Por tanto $|\psi_1\rangle \ R \ |\psi_3\rangle$.

- R es reflexiva: $|\psi\rangle = 1|\psi\rangle = e^{i0}|\psi\rangle$, luego $|\psi\rangle \ R \ |\psi\rangle$.

- R es simétrica: supongamos $|\psi_1\rangle \ R \ |\psi_2\rangle$, entonces existe ϕ tal que $e^{i\phi}|\psi_1\rangle = |\psi_2\rangle$, pasando $e^{i\phi}$ al otro lado obtenemos $e^{-i\phi}|\psi_2\rangle = |\psi_1\rangle$ y nos permite concluir $|\psi_2\rangle \ R \ |\psi_1\rangle$.

□

Dado $|\psi\rangle$, si consideramos su clase de equivalencia $[|\psi\rangle]_R$ y $|\xi\rangle \in [|\psi\rangle]_R$, las probabilidades de medir $|0\rangle$ y de medir $|1\rangle$ para $|\psi\rangle$ y $|\xi\rangle$ son la iguales. Pero el recíproco no es cierto. Por ejemplo, sean:

$$|\psi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$$
$$|\xi\rangle = \frac{i}{\sqrt{2}}|0\rangle - \frac{i}{\sqrt{2}}|1\rangle .$$

Ambos qubits tienen la misma probabilidad para medir $|0\rangle$ y $|1\rangle$, $|\frac{1}{\sqrt{2}}|^2 = |\frac{i}{\sqrt{2}}|^2 = |\frac{-i}{\sqrt{2}}|^2$, pero no están relacionados por R.

Pasando los coeficientes a polares obtenemos

$$|\psi\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$$
$$|\xi\rangle = \frac{1}{\sqrt{2}}e^{\frac{i\pi}{2}}|0\rangle + \frac{1}{\sqrt{2}}e^{-\frac{i\pi}{2}}|1\rangle .$$

Si suponemos, $|\psi\rangle \ R \ |\xi\rangle$, existiría $e^{i\phi}$ tal que

$$|\psi\rangle = e^{i\phi}\,|\xi\rangle$$
$$\frac{1}{\sqrt{2}}\,|0\rangle + \frac{1}{\sqrt{2}}\,|1\rangle = \frac{1}{\sqrt{2}}e^{i(\phi+\frac{\pi}{2})}\,|0\rangle + \frac{1}{\sqrt{2}}e^{i(\phi-\frac{\pi}{2})}\,|1\rangle\,.$$

Por ser $|0\rangle$ y $|1\rangle$ una base de $\mathbb{C}^2$, tiene que ocurrir igualdad coeficiente a coeficiente.

$$\frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}e^{i(\phi+\frac{\pi}{2})} \implies 1 = e^{i(\phi+\frac{\pi}{2})} \implies \phi+\frac{\pi}{2} = 2k\pi\ ,\ k\in\mathbb{Z} \implies \phi = \frac{(4k-1)\pi}{2}\ ,\ k\in\mathbb{Z},$$

$$\frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}e^{i(\phi-\frac{\pi}{2})} \implies 1 = e^{i(\phi-\frac{\pi}{2})} \implies \phi-\frac{\pi}{2} = 2h\pi\ ,\ h\in\mathbb{Z} \implies \phi = \frac{(4h+1)\pi}{2}\ ,\ h\in\mathbb{Z}.$$

Juntando ambas conclusiones

$$\frac{(4k-1)\pi}{2} = \frac{(4h+1)\pi}{2},\ k,h\in\mathbb{Z} \implies 4(k-h) = 2\ ,\ k,h\in\mathbb{Z}.$$

Llegamos a una contradicción porque 2 no es multiplo entero de 4. Luego $|\psi\rangle$ y $|\xi\rangle$ no están relacionados por R. Equivalentemente, $|\xi\rangle \notin [|\psi\rangle]_R$.

En el anterior ejemplo pasamos a polares los qubits para probar la contradicción. Igualmente, en general, dado un $|\psi\rangle = a\,|0\rangle + b\,|1\rangle$, podemos pasar los coeficientes a y b a forma polar obteniendo

$$|\psi\rangle = r_1e^{i\phi_1}\,|0\rangle + r_2e^{i\phi_2}\,|1\rangle\,.$$

Donde $r_1 = |a|$ y $r_2 = |b|$, luego $r_1 \geq 0$. Además $r_1^2 + r_2^2 = 1$, por tanto $r_1 \leq 1$. Definimos $\theta = 2\arccos(r_1) \in [0,\pi]$, luego $r_1 = \cos(\frac{\theta}{2})$. Siguiendo la misma línea $r_1^2 + r_2^2 = 1 = \cos^2(\frac{\theta}{2}) + \sin^2(\frac{\theta}{2})$, entonces $r_2 = \pm\sin(\frac{\theta}{2})$ y como $r_2 \geq 0, \theta \in [0,\pi]$, concluimos $r_2 = \sin(\frac{\theta}{2})$. Reescribimos $|\psi\rangle$,

$$|\psi\rangle = \cos(\frac{\theta}{2})e^{i\phi_1}\,|0\rangle + \sin(\frac{\theta}{2})e^{i\phi_2}\,|1\rangle\,.$$

Si ahora multiplicamos por $e^{-i\phi_1}$, nos quedaría

$$|\psi\rangle_R = \cos(\frac{\theta}{2})\,|0\rangle + \sin(\frac{\theta}{2})e^{i(\phi_2-\phi_1)}\,|1\rangle\,.$$

Trivialmente $|\psi\rangle_R \in [|\psi\rangle]_R$ y $|\psi\rangle_R$ será por defecto el representante de está clase. El valor $\phi_2 - \phi_1$ se denomina cambio de fase.

Ahora estamos en condiciones de definir la esfera de Bloch, sea $\mathbf{S}^2 = \{(x,y,z) \in \mathbb{R} : x^2+y^2+z^2 = 1\}$. Consideramos la aplicación $\Phi : \mathcal{Q}/R \longrightarrow \mathbf{S}^2$

$$\Phi([\cos(\frac{\theta}{2})e^{i\phi_1}\,|0\rangle + \sin(\frac{\theta}{2})e^{i\phi_2}\,|1\rangle]_R) = (\sin(\theta)\cos(\phi_2-\phi_1), \sin(\theta)\sin(\phi_2-\phi_1), \cos(\theta)).$$

Proposición 1.3. *La aplicación Φ está bien definida y es biyectiva.*

Demostración. Primero probaremos que está bien definida, después, que es inyectiva y finalmente sobreyectiva.

- Bien definida: dado $|\psi\rangle$ sean $|\xi_1\rangle, |\xi_2\rangle \in [|\psi\rangle]_R$, por tanto $|\xi_2\rangle = e^{\phi_3}|\xi_1\rangle$. Si escribimos en polares $|\xi_1\rangle = \cos(\frac{\theta}{2})e^{i\phi_1}|0\rangle + \sin(\frac{\theta}{2})e^{i\phi_2}|1\rangle$ entonces $|\xi_2\rangle = \cos(\frac{\theta}{2})e^{i(\phi_1+\phi_3)}|0\rangle + \sin(\frac{\theta}{2})e^{i(\phi_2+\phi_3)}|1\rangle$.

  Φ estará bien definida si $\Phi([|\xi_1\rangle]_R) = \Phi([|\xi_2\rangle]_R)$, desarrollando componente a componente:

$$\begin{aligned}\sin(\theta)\cos(\phi_2-\phi_1) &= \sin(\theta)\cos(\phi_2+\phi_3-\phi_1-\phi_3)\\ \sin(\theta)\sin(\phi_2-\phi_1) &= \sin(\theta)\sin(\phi_2+\phi_3-\phi_1-\phi_3)\\ \cos(\theta) &= \cos(\theta).\end{aligned}$$

  Claramente se verfica componente a componente, así que podemos concluir que Φ está bien definida.

- Inyectividad: supongamos $\Phi([|\psi\rangle]_R) = \Phi([|\xi\rangle]_R)$ y queremos ver $[|\psi\rangle]_R = [|\xi\rangle]_R$. Como Φ está bien definida, puedo suponer, $|\psi\rangle = \cos(\frac{\theta_\psi}{2})|0\rangle + \sin(\frac{\theta_\psi}{2})e^{i\phi_\psi}|1\rangle$,$|\xi\rangle = \cos(\frac{\theta_\xi}{2})|0\rangle + \sin(\frac{\theta_\xi}{2})e^{i\phi_\xi}|1\rangle$ con $\theta_\psi, \theta_\xi \in [0,\pi]$, si miramos componente a componente la igualdad $\Phi([|\psi\rangle]_R) = \Phi([|\xi\rangle]_R)$ obtenemos

$$\begin{aligned}\sin(\theta_\phi)\cos(\phi_\psi) &= \sin(\theta_\xi)\cos(\phi_\xi)\\ \sin(\theta_\phi)\sin(\phi_\psi) &= \sin(\theta_\xi)\sin(\phi_\xi)\\ \cos(\theta_\phi) &= \cos(\theta_\xi).\end{aligned}$$

  Observando la tercera igualdad y teniendo en cuenta $\theta_\phi, \theta_\xi \in [0,\pi]$, aplicamos arcocoseno en ambos lados, resultando en $\theta_\phi = \theta_\xi$.

  Si $\theta_\phi \neq 0, \pi$ entonces $\sin(\theta_\phi) = \sin(\theta_\xi) \neq 0$, así que puede multiplicar por su inverso en las dos primeras igualdades quedando

$$\begin{aligned}\cos(\phi_\psi) &= \cos(\phi_\xi)\\ \sin(\phi_\psi) &= \sin(\phi_\xi).\end{aligned}$$

  Estas dos igualdades implican $2k\pi + \phi_\psi = \phi_\xi$ para algún $k \in \mathbb{Z}$. Incorporando estas igualdades en los qubits nos queda

$$\cos(\frac{\theta_\psi}{2})|0\rangle + \sin(\frac{\theta_\psi}{2})e^{i\pi\phi_\psi}|1\rangle = \cos(\frac{\theta_\psi}{2})|0\rangle + \sin(\frac{\theta_\psi}{2})e^{i2k\pi+i\phi_\psi}|1\rangle.$$

  Como $e^{i2k\pi} = 1$ con $k \in \mathbb{Z}$, los representantes escogidos son el mismo, por tanto $[|\psi\rangle]_R = [|\xi\rangle]_R$.

  Si $\theta_\phi = 0$, entonces $\sin(\frac{\theta_\phi}{2}) = \sin(\frac{\theta_\xi}{2}) = \sin(0) = 0$. Mirando los qubits tenemos $|\phi\rangle = |0\rangle = |\xi\rangle$ y concluimos $[|\psi\rangle]_R = [|\xi\rangle]_R$.

Si $\theta_\phi = \pi$, entonces $\cos(\frac{\theta_\phi}{2}) = \cos(\frac{\theta_\xi}{2}) = \cos(\frac{\pi}{2}) = 0$. Volviendo a examinar los qubits, obtenemos $|\psi\rangle = e^{i\phi_\psi}|1\rangle$ y $|\xi\rangle = e^{i\phi_\xi}|1\rangle$. Si multiplicamos $e^{i(-\phi_\xi+\phi_\psi)}|\xi\rangle = |\phi\rangle \implies [|\psi\rangle]_R = [|\xi\rangle]_R$.

- Sobreyectividad: dado $(x,y,z) \in \mathbf{S}^2$ sabemos que existen $\theta, \phi \in \mathbb{R}$ tales que: $(x,y,z) = (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))$. Sea el qubit $|\psi\rangle = \cos(\frac{\theta}{2})|0\rangle + \sin(\frac{\theta}{2})e^{i\phi}|1\rangle$, entonces, por la propia definición de Φ, tenemos $\Phi([|\psi\rangle]_R) = (x,y,z)$.

□

Ya tenemos propiamente definida la esfera de Bloch, y por el fin de ahorrar notación, se dejará de escribir $\Phi([|\psi\rangle]_R)$, sustituyéndolo por $\Phi([|\psi\rangle])$. Pero antes de ver algunos ejemplos de qubits proyectados vamos a presentar dos pares de qubits especialmente interesantes:

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}\ ,\ |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}} \qquad (1.1)$$
$$|i\rangle = \frac{|0\rangle + i|1\rangle}{\sqrt{2}}\ ,\ |-i\rangle = \frac{|0\rangle - i|1\rangle}{\sqrt{2}}.$$

Si comprobamos la ortogonalidad

$$\begin{aligned}
\langle -||+\rangle &= (\frac{\langle 0| - \langle 1|}{\sqrt{2}})(\frac{|0\rangle + |1\rangle}{\sqrt{2}}) \\
&= \frac{\langle 0||0\rangle + \langle 0||1\rangle - \langle 1||0\rangle - \langle 1||1\rangle}{2} \\
&= \frac{\langle 0||0\rangle - \langle 1||1\rangle}{2} = \frac{1}{2} - \frac{1}{2} = 0 \\
\langle -i||i\rangle &= (\frac{\langle 0| + i\langle 1|}{\sqrt{2}})(\frac{|0\rangle + i|1\rangle}{\sqrt{2}}) \\
&= \frac{\langle 0||0\rangle + i\langle 0||1\rangle + i\langle 1||0\rangle - \langle 1||1\rangle}{2} \\
&= \frac{\langle 0||0\rangle - \langle 1||1\rangle}{2} = \frac{1}{2} - \frac{1}{2} = 0.
\end{aligned}$$

Como $\{|+\rangle, |-\rangle\}$ son ortogonales entre si forman una base de $\mathbb{C}^2$, y por el mismo motivo $\{|+i\rangle, |-i\rangle\}$ también es una base.

Ahora veamos estos qubits, además de los ya conocidos $|0\rangle, |1\rangle$, en la esfera. Se puede ver una representanción gráfica en 1.1:

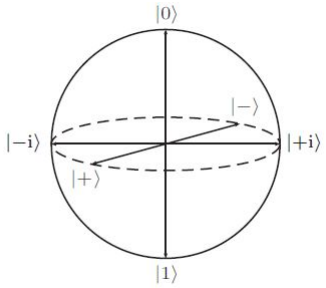


Figura 1.1.: Imagen de la esfera de Bloch tomada de https://stem.mitre.org/quantum/quantum-concepts/bloch-sphere.html

$$\Phi(|0\rangle) = \Phi(\cos(0)\,|0\rangle) = (0,0,1)$$
$$\Phi(|1\rangle) = \Phi(\sin(\frac{\pi}{2})\,|1\rangle) = (0,0,1)$$
$$\Phi(|+\rangle) = \Phi(\cos(\frac{\pi}{4})\,|0\rangle + \sin(\frac{\pi}{4})\,|1\rangle) = (1,0,0)$$
$$\Phi(|-\rangle) = \Phi(\cos(\frac{\pi}{4})\,|0\rangle + \sin(\frac{\pi}{4})e^{i\pi}\,|1\rangle) = (-1,0,0)$$
$$\Phi(|+i\rangle) = \Phi(\cos(\frac{\pi}{4})\,|0\rangle + \sin(\frac{\pi}{4})e^{i\frac{\pi}{2}}\,|1\rangle) = (0,1,0)$$
$$\Phi(|-i\rangle) = \Phi(\cos(\frac{\pi}{4})\,|0\rangle + \sin(\frac{\pi}{4})e^{-i\frac{\pi}{2}}\,|1\rangle) = (0,-1,0).$$

En primer lugar vemos que $\{|0\rangle, |1\rangle\}$,$\{|+\rangle, |-\rangle\}$,$\{|+i\rangle, |-i\rangle\}$ se proyectan en los ejes Z,X,Y respectivamente. Aún más llamativo es el hecho de que han ido a puntos antípodas. Esta propiedad la recogemos con la siguiente proposición.

Proposición 1.4. *Sean $|\psi\rangle$ y $|\xi\rangle$ dos qubits. Entonces $\langle\psi|\,|\xi\rangle = 0$ si, y solo si, $\Phi(|\psi\rangle) = -\Phi(|\xi\rangle)$.*

Demostración. Podemos suponer $|\psi\rangle = \cos(\frac{\theta_\phi}{2})\,|0\rangle + \sin(\frac{\theta_\phi}{2})e^{i\phi_\psi}\,|1\rangle$ y $|\xi\rangle = \cos(\frac{\theta_\xi}{2})\,|0\rangle + \sin(\frac{\theta_\xi}{2})e^{i\phi_\xi}\,|1\rangle$.

Si los escribimos en polares, $|\psi\rangle = \cos(\frac{\theta_1}{2})e^{i\phi_{11}}\,|0\rangle + \sin(\frac{\theta_1}{2})e^{i\phi_{12}}\,|1\rangle$, $|\xi\rangle = \cos(\frac{\theta_2}{2})e^{i\phi_{21}}\,|0\rangle + \sin(\frac{\theta_{22}}{2})e^{i\phi_2}\,|1\rangle$, como $\langle\psi|\,|\xi\rangle = 0$ entonces $\overline{e^{-i\phi_{11}}}e^{-i\phi_{21}}\,\langle\psi|\,|\xi\rangle = 0$ que es lo mismo que decir que $= \cos(\frac{\theta_1}{2})\,|0\rangle + \sin(\frac{\theta_1}{2})e^{i\phi_{12}-i\phi_{11}}\,|1\rangle$ y $\cos(\frac{\theta_2}{2})\,|0\rangle + \sin(\frac{\theta_2}{2})e^{i\phi_{22}-i\phi_{21}}\,|1\rangle$ son ortogonales. Para recuperar mis estados originales me basta con multiplicar cada qubit por la unidad correspondiente.

De izquierda a derecha, sabemos $\langle\psi|\,|\xi\rangle = 0$

$$\begin{aligned}\langle\psi|\,|\xi\rangle &= \begin{pmatrix}\cos(\frac{\theta_\psi}{2}) & \sin(\frac{\theta_\psi}{2})e^{-i\phi_\psi}\end{pmatrix}\begin{pmatrix}\cos(\frac{\theta_\xi}{2})\\ \sin(\frac{\theta_\xi}{2})e^{i\phi_\xi}\end{pmatrix}\\ &= \cos(\tfrac{\theta_\psi}{2})\cos(\tfrac{\theta_\xi}{2}) + \sin(\tfrac{\theta_\psi}{2})\sin(\tfrac{\theta_\xi}{2})e^{i(\phi_\xi-\phi_\psi)} = 0 \implies\\ \cos(\tfrac{\theta_\psi}{2})\cos(\tfrac{\theta_\xi}{2}) &= -\sin(\tfrac{\theta_\psi}{2})\sin(\tfrac{\theta_\xi}{2})e^{i(\phi_\xi-\phi_\psi)} = \sin(\tfrac{\theta_\psi}{2})\sin(\tfrac{\theta_\xi}{2})e^{i(\phi_\xi-\phi_\psi+\pi)}.\end{aligned}$$

Dos complejos son iguales en forma polar si, y solo si, los modulos son iguales y los exponentes difieren por un multiplo de 2π son el mismo. Incorporándolo a la última igualdad nos queda

$$\cos(\tfrac{\theta_\psi}{2})\cos(\tfrac{\theta_\xi}{2}) = \sin(\tfrac{\theta_\psi}{2})\sin(\tfrac{\theta_\xi}{2}) \implies \cos(\tfrac{\theta_\psi}{2})\cos(\tfrac{\theta_\xi}{2}) - \sin(\tfrac{\theta_\psi}{2})\sin(\tfrac{\theta_\xi}{2}) = \cos(\frac{\theta_\psi+\theta_\xi}{2}) = 0$$

$$\implies \frac{\theta_\psi+\theta_\xi}{2} = (2h+1)\frac{\pi}{2}\ ,\ h\in\mathbb{Z} \implies \theta_\xi = (2h+1)\pi - \theta_\psi\ ,\ h\in\mathbb{Z}$$
$$\phi_\xi - \phi_\psi + \pi = 2k\pi\ ,\ k\in\mathbb{Z} \implies \phi_\xi = \phi_\psi + (2k-1)\pi\ ,\ k\in\mathbb{Z}.$$

En está representación de los qubits imponemos $\theta_\psi, \theta_\xi \in [0,\pi]$, necesariamente $h = 0$ y por tanto $\phi_\xi = \pi - \theta_\psi$. Relacionados $|\phi\rangle$ y $|\xi\rangle$, pasemos a comprobar que sus correspondientes en la esfera de Bloch son antípodas,

$$\begin{aligned}\cos(\theta_\xi) &= \cos(\pi-\theta_\psi) = -\cos(\theta_\psi)\\ \sin(\theta_\xi) &= \sin(\pi-\theta_\psi) = \sin(\theta_\psi)\\ \cos(\phi_\xi) &= \cos(\phi_\psi + (2k-1)\pi) = \cos(\phi_\psi-\pi) = -\cos(\phi_\psi)\\ \sin(\phi_\xi) &= \sin(\phi_\psi + (2k-1)\pi) = \sin(\phi_\psi-\pi) = -\sin(\phi_\psi).\end{aligned} \tag{1.2}$$

A la luz de las igualdades anteriores y de la definición de Φ, es inmediato deducir, $\Phi(|\psi\rangle) = -\Phi(|\xi\rangle)$.

Probemos la implicación de derecha a izquierda, suponemos $\Phi(|\psi\rangle) = -\Phi(|\xi\rangle)$, desglosando coordenada a coordenada:

$$\begin{aligned}\sin(\theta_\psi)\cos(\phi_\psi) &= -\sin(\theta_\xi)\cos(\phi_\xi)\\ \sin(\theta_\psi)\sin(\phi_\psi) &= -\sin(\theta_\xi)\sin(\phi_\xi)\\ \cos(\theta_\psi) &= -\cos(\theta_\xi).\end{aligned}$$

Sabiendo qué $\theta_\psi, \theta_\xi \in [0,\pi]$ y repitiendo un proceso análogo al hecho en (1.2), concluimos,

$\theta_{\xi} = \pi - \theta_{\psi}$ y $\theta_{\xi} = \phi_{\psi} + (2k-1)\pi$. Veamos que $|\psi\rangle$ y $|\xi\rangle$ son ortogonales.

$$\begin{aligned}\langle\psi|\,|\xi\rangle &= \cos(\frac{\theta_{\psi}}{2})\cos(\frac{\theta_{\xi}}{2}) + \sin(\frac{\theta_{\psi}}{2})\sin(\frac{\theta_{\xi}}{2})e^{i(\phi_{\xi}-\phi_{\psi})} \\ &= \cos(\frac{\theta_{\psi}}{2})\cos(\frac{\pi-\theta_{\psi}}{2}) + \sin(\frac{\theta_{\psi}}{2})\sin(\frac{\pi-\theta_{\psi}}{2})e^{i(\phi_{\psi}-\phi_{\psi}+(2k-1)\pi)} \\ &= \cos(\frac{\theta_{\psi}}{2})\cos(\frac{\pi-\theta_{\psi}}{2}) - \sin(\frac{\theta_{\psi}}{2})\sin(\frac{\pi-\theta_{\psi}}{2}) \\ &= \cos(\frac{\pi-\theta_{\psi}+\theta_{\psi}}{2}) = \cos(\frac{\pi}{2}) = 0.\end{aligned}$$

□

1.3. Usando varios qubits

Al igual que con los bits, es necesario ser capaces de trabajar con más de un qubit a la vez dentro del mismo sistema y que estos puedan interactuar. La operación matemática que permite operar con varios qubits es el producto tensor.

El producto tensor se define a partir de dos espacios vectoriales, V y W, sobre un mismo cuerpo, K. Consideramos el espacio vectorial que tiene como base a los elementos de $V \times W$, dicho espacio sería $F(V \times W) = \{\sum_{i=1}^{n} c_i(v_i, w_i) : c_i \in K, (v_i, w_i) \in V \times W, 1 \leq i \leq n, n \in \mathbb{N}\}$ donde (v_i, w_i) son linealmente independidientes si $(v_i, w_i) \neq (v_j, w_j)$. Una forma de identificar al espacio vectorial $F(V \times W)$ son a través de las funciones de $f : V \times W \to \mathbb{R}$ que tienen un número finito de 0.

Si tenemos una f que solamente no se anula en $(v_1, w_1), \ldots, (v_n, w_n)$ y cuyos valores en dichos puntos son $f(v_i, w_i) = c_i$. Entonces se puede identificar a f en $F(V \times W)$ como el elemento $\sum_{i=1}^{n} c_i(v_i, w_i)$.

Teniendo una idea del espacio vectorial $F(F(V \times W))$, definimos en él las siguientes relaciones de equivalencia:

1. $(v_1 + v_2, w) \sim (v_1, w) + (v_2, w)$, $\forall v_1, v_2 \in V$, $w \in W$.
2. $(v, w_1 + w_2) \sim (v, w_1) + (v, w_2)$, $\forall v \in V$, $w_1, w_2 \in W$.
3. $(cv, w) \sim c(v, w)$, $\forall c \in K$, $v \in V$, $\forall w \in W$.
4. $(v, cw) \sim c(v, w)$, $\forall c \in K$, $v \in V$, $\forall w \in W$.

Sea R al espacio generado por las cuatro relaciones de equivalencia anteriores.

Definición 1.2. Definimos al producto tensor entre V y W espacios vectoriales sobre el mismo cuerpo K, denotado por $V \otimes W$, como

$$V \otimes W = F(V \times W)/R.$$

Dado $(v, w) \in V \times W$ se denomina tensor a la clase de equivalencia de (v, w) y se denota $v \otimes w$.

Las relaciones de equivalencias anteriores en $V \otimes W$ se transforman en igualdades, también se verifica que $V \otimes W$ es un espacio vectorial. Dadas B_V y B_W bases de V y W respectivamente, entonces una base de $V \otimes W$ es

$$B = \{v \otimes w : v \in B_V, w \in B_W\}.$$

Luego si V y W tienen dimensión n y m respectivamente, entonces $V \otimes W$ tiene dimensión nm.

Si además V y W son espacios de Hilbert, $V \otimes W$ es un espacio de Hilbert. El producto escalar de dicho espacio sobre los tensores $v \otimes w$ es

$$< v_1 \otimes w_1, v_2 \otimes w_2 >=< v_1, v_2 >< w_1, w_2 > , \ \forall v \in V, w \in W$$

extendiendo por linealiadad y antilinealidad al total del espacio.

Centrándonos en qubits, cuando trabajamos con n qubits trabajamos en el producto tensorial de cada espacio donde vive cada qubit. Cada qubit vive en $\mathbb{C}^2$, luego, al trabajar con n qubits, el espacio natural donde trabajamos es $\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2 = \mathbb{C}^{2^n}$.

Sean $|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_n\rangle \in \mathbb{C}^2$ qubits, el sistema formado por estos qbits está formado por el elemento en el producto tensor

$$|\psi_1\rangle \otimes |\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle = |\psi_1\rangle |\psi_2\rangle \ldots |\psi_n\rangle = |\psi_1 \psi_2 \ldots \psi_n\rangle \in \mathbb{C}^{2^n}. \tag{1.3}$$

Las igualdades (1.3) representan diversas formas para notar varios qubits, la más usual es la que aparece en la última igualdad.

Varios ejemplos son:

$$|00\rangle, \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \frac{i\,|00\rangle - |10\rangle + e^{i\frac{\pi}{4}}\,|11\rangle}{\sqrt{3}}.$$

Calculemos la norma de algunos de los ejemplos anteriores:

$$\begin{aligned}
< |00\rangle, |00\rangle > &=< |0\rangle \otimes |0\rangle, |0\rangle \otimes |0\rangle >=< |0\rangle, |0\rangle >< |0\rangle, |0\rangle >= 1 \\
< \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \frac{|00\rangle + |11\rangle}{\sqrt{2}} > &= \frac{1}{2} < |00\rangle + |11\rangle, |00\rangle + |11\rangle > \\
&= \frac{1}{2}(< |00\rangle, |00\rangle > + < |00\rangle, |11\rangle > \\
&+ < |11\rangle, |00\rangle > + < |11\rangle, |11\rangle >) \\
&= \frac{1}{2}(< |0\rangle, |0\rangle >< |0\rangle, |0\rangle > + < |0\rangle, |1\rangle >< |0\rangle, |1\rangle > \\
&+ < |1\rangle, |0\rangle >< |1\rangle, |0\rangle > + < |1\rangle, |1\rangle >< |1\rangle, |1\rangle >) = 1.
\end{aligned}$$

Vemos que la norma de los anteriores vectores es 1, ya que su norma es 1. De hecho $\||\psi\xi\rangle\| = \||\psi\rangle\|\||\xi\rangle\| = 1$.

Por tanto es natural exigir que si tenemos un estado cuántico resultante de combinar varios qubits, este tenga norma 1.

Como último comentario respecto a como se agrupan los qubits, la forma en la que se agrupan n qubits da un espacio de estados de dimensión 2^n. Este crecimiento exponencial sugiere una posible aceleración exponencial de la computación en ordenadores cuánticos sobre ordenadores clásicos. Esto es una diferencia fundamental entre bits y qubits porque con bits el crecimiento en el espacio de estados es de $2n$ y la aceleración es lineal.

1.4. Entrelazamiento

Definición 1.3. Sea $|\psi\rangle \in \mathbb{C}^{2^n} \otimes \mathbb{C}^{2^m}$ un estado cuántico formado por varios qubits, decimos que $|\psi\rangle$ es un estado entrelazado si no se puede poner como tensor de dos estados. En otro caso diremos que el estado $|\psi\rangle$ es separable.

Consideremos el sistema formado por dos qubits cuyo estado corresponde con

$$\frac{|01\rangle + |10\rangle}{\sqrt{2}} \tag{1.4}$$

Veamos que los qubits están entrelazados. Para ello supongamos que no lo están, es decir, podemos descomponerlo como producto tensor de otros dos qubits:

$$\begin{aligned} \frac{|01\rangle + |10\rangle}{\sqrt{2}} &= |\psi\rangle \otimes |\xi\rangle = (a\,|0\rangle + b\,|1\rangle) \otimes (c\,|0\rangle + d\,|1\rangle) \\ &= ac\,|0\rangle \otimes |0\rangle + ad\,|0\rangle \otimes |1\rangle + bc\,|1\rangle \otimes |0\rangle + bd\,|1\rangle \otimes |1\rangle \\ &= ac\,|00\rangle + ad\,|01\rangle + bc\,|10\rangle + bd\,|11\rangle . \end{aligned} \tag{1.5}$$

Como $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ es una base de $\mathbb{C}^2 \otimes \mathbb{C}^2$, para poder verificarse la igualdad (1.5) tiene que cumplirse:

$$\begin{aligned} ac = 0 &\implies a = 0 \vee c = 0 \\ ad = \frac{1}{\sqrt{2}} &\implies a \neq 0 \neq d \\ bc = \frac{1}{\sqrt{2}} &\implies b \neq 0 \neq c \\ bd = 0 &\implies b = 0 \vee d = 0. \end{aligned} \tag{1.6}$$

Las implicaciones de (1.6) dan una contradicción. Análogamente se prueba que los siguientes estados también están entrelazados:

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}}\,,\ \frac{|01\rangle - |10\rangle}{\sqrt{2}}\,,\ \frac{|00\rangle + |11\rangle}{\sqrt{2}}\,,\ \frac{|01\rangle + |10\rangle}{\sqrt{2}}.$$

De los anteriores destacamos al primero, que denominaremos par ERP.

Teorema 1.2 (Descomposición de Schmidt)**.** *Sea $|\psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^n$ de módulo uno, entonces existen dos bases ortonormales $\{u_1 \ldots, u_n\}$,$\{v_1 \ldots, v_n\}$ de $\mathbb{C}^n$ tal que*

$$|\psi\rangle = \sum_{i=1}^{n} c_i\,|u_i v_i\rangle\,.$$

Con c_i reales no negativos verificando $\sum_{i=1}^{n} c_i^2 = 1$.

Definición 1.4. Se define como número de Schmidt al número de $|\psi\rangle$ al número de c_i no nulos en la descomposición

$$|\psi\rangle = \sum_{i=1}^{n} c_i \, |u_i v_i\rangle \, .$$

Observación 1.1. $|\psi\rangle$ es separable si, y solo si, su número de Schmidt es 1.

2. Operadores

En el anterior capítulo describimos la unidad de información básica en un ordenador cuántico, el qubit. Esta unidad representa el primer postulado de la teoría de la computación cuántica. Existen diversos postulados dependiendo de donde se consulte, nosotros vamos a enunciar 4 postulados que son con los que trabajaremos.

1. **El estado del sistema.**

   El estado de un sistema cuántico es un vector $|\psi(t)\rangle$ que varía en el tiempo en un espacio de Hilbert. Dicho estado contiene toda la información que podemos obtener en el sistema. Trabajamos con estados normalizados $\langle\psi(t)|\psi(t)\rangle = 1$.

   Para el desarrollo realizado, tomamos esto vectores como qubits los cuales ya describimos en el anterior capítulo.

2. **Cantidades observables corresponde a operadores.**

   Toda variable dinámica A que sea físicamente medible corresponde a un operador lineal hermítico A.

   En el contexto de computación cuántica cuando se mencionan operadores se sobreentienden que son operadores lineales hermíticos.

3. **Mediciones.**

   Las posibles mediciones de una variable dinámica A son los valores propios a_n del operador A que corresponde a dicha variable.

4. **Evolución en el tiempo del sistema.** La evolución de un sistema cuántico aislado es gobernada por la ecuación de Schrödinger

   $$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle$$

   donde H es un operador denominado Hamiltoniano del sistema. Este operador corresponde a la energía total del sistema. Los posibles niveles de energía que puede tener el sistema corresponden a los valores propios de H. La evolución en el tiempo del vector de estados es

   $$|\psi(t)\rangle = e^{-iHt/\hbar}|\psi(0)\rangle.$$

   Luego la evolución en el tiempo de un estado cuántico está gobernado por el operador

   $$U = e^{-iHt/\hbar}.$$

En este capítulo nos centraremos en el segundo y tercer postulado. El cuarto postulado no lo trataremos en profundidad pero dada su relevancia se incluye.

2.1. Preeliminares de operadores

Con el fin de fijar notación y tener presentes propiedades de los operadores, esta primera sección del capítulo constará de un repaso de los operadores linales y matrices.

Definición 2.1. Sea $\hat{A} : V_1 \to V_2$ un operador entre espacios vectoriales complejos. Diremos que $\hat{A}$ es un operador lineal o aplicación lineal si verifica

$$\hat{A}(av + bw) = a\hat{v} + b\hat{w}, \forall a, b \in \mathbb{C}, v, w \in V_1.$$

Si tengo un operador lineal,$\hat{A}$, entre dos espacios vectoriales complejos V y W cada uno de dimensión $\dim(V) = n$ y $\dim(W) = m$ y dadas $B_V = \{v_1, \ldots, v_n\}$ y $B_W = \{w_1, \ldots, w_m\}$ bases de V y W. Entonces existe una matriz en $\mathbb{C}^{m\times n}$ que representa la acción de la aplicación desde la base B_V en la base B_W, a esta matriz la notaremos

$$M(\hat{A}, B_W \leftarrow B_V).$$

Dadas otras dos bases B'_V y B'_W, existe una relación entre $M(\hat{A}, B_W \leftarrow B_V)$ y $M(\hat{A}, B_W \leftarrow B_V)$ siendo esta

$$M(\hat{A}, B'_W \leftarrow B'_V) = P_W M(\hat{A}, B_W \leftarrow B_V) P_V$$

donde $P_V \in \mathbb{C}^{n\times n}$ y $P_W \in \mathbb{C}^{m\times m}$ son

$$P_V = M(I, B_V \leftarrow B'_V)\ ,\ P_W = M(I, B'_W \leftarrow B_W)$$

con I la aplicación identidad. Además P_V y P_W son ambas invertibles, con inversa

$$P_V^{-1} = M(I, B'_V \leftarrow B_V)\ ,\ P_W^{-1} = M(I, B_W \leftarrow B'_W).$$

Si $V = W$ y $B_V = B_W$ reduciremos la notación a

$$M(\hat{A}, B_V \leftarrow B_V) = M(\hat{A}, B_V).$$

Definición 2.2. Dadas $\hat{A} : V_1 \to W_1$ y $\hat{B} : V_2 \to W_2$ son dos aplicaciones lineales. Definimos la aplicación $\hat{A} \otimes \hat{B} : V_1 \otimes V_2 \to W_1 \otimes W_2$ a aquella que extiende por linealidad la regla

$$\hat{A} \otimes \hat{B}(v_1 \otimes v_2) = \hat{A}(v_1) \otimes \hat{B}(v_2)\ ,\ \forall v_1 \in V_1\ ,\ v_2 \in V_2.$$

Esta aplicación es el producto tensorial entre $\hat{A}$ y $\hat{B}$.

Definición 2.3. Sea V espacio de hilbert de dimension n, definimos la traza de una aplicación lineal $\hat{A}$ como la traza de $M(\hat{A}, B)$ donde B es una base ortonormal de V. La denotaremos como $\mathrm{tr}(\hat{A}) = \mathrm{tr}(M(\hat{A}, B))$.

La traza no depende de la base escogida para representar al operador $\hat{A}$. Una forma práctica de calcular la traza de $\hat{A}$ dada una base ortonormal $B = \{u_1, \ldots, u_n\}$ es

$$\mathrm{tr}(\hat{A}) = \sum_{i=0}^{n} < u_i | \hat{A}(u_i) > .$$

Definición 2.4. Sea $\hat{A}$ una aplicación lineal de V en V. Decimos que λ es un valor propio de

A si existe un vector no nulo $v \in V$ tal que

$$\hat{A}(v) = \lambda v.$$

En cuyo caso diremos que v es un vector propio asociado al valor propio λ.

Analogámente, sea $M \in \mathbb{C}^{n\times n}$. Decimos que λ es un valor propio de M si existe un vector no nulo $v \in \mathbb{C}^n$ tal que

$$Mv = \lambda v.$$

En cuyo caso diremos que v es un vector propio asociado al valor propio λ.

Definición 2.5. Sea $\hat{A}$ una aplicación lineal de V en V. Diremos que $\hat{A}$ es diagonalizable si existe una base ortogonal B de V tal que $M(\hat{A}, B)$ es una matriz diagonal.

Proposición 2.1. *Supongamos $\hat{A}$ diagonalizable, por tanto existe B base ortogonal tal que $M(\hat{A}, B)$ es una matriz diagonal. Sean $\lambda_1, \ldots, \lambda_n$ los valores de la diagonal $M(\hat{A}, B)$, entonces dichos valores son valores propios de $\hat{A}$.*

Definición 2.6. Dada una $M \in \mathbb{C}^{n\times n}$ se define su adjunto hermítico como:

$$M^\dagger = \overline{M^t}.$$

Definición 2.7. Dado un operador lineal $A : \mathbb{C}^n \to \mathbb{C}^n$, si para cada $y \in \mathbb{C}^n$ se verifica

$$< x|A(y) >=< z|y >$$

para algún $z \in \mathbb{C}^n$. Entonces se dice que x está en el dominio del operador adjunto de A, $A^\dagger$, y se define su imagen como

$$A^\dagger(x) = z.$$

Proposición 2.2. *Si $A : \mathbb{C}^n \to \mathbb{C}^n$ un operador lineal y B una base ortonormal de $\mathbb{C}^n$. Entonces $M(A^\dagger, B) = M(A, B)^\dagger$.*

Por ejemplo, sea el operador $\hat{A}$ definido por la siguiente matriz en una base ortonormal

$$\begin{pmatrix} i & 4 \\ 5-6i & 0 \end{pmatrix}$$

entonces, la expresión matricial del adjunto hermítico de $\hat{A}$ es

$$\begin{pmatrix} -i & 5+6i \\ 4 & 0 \end{pmatrix}.$$

Definición 2.8. Sea A un operador lineal de V en V, $\dim(V) = n$.

- A es hermítico si A=$A^\dagger$. Suponemos que todo observable es hermítico.
- A es unitaria si $AA^\dagger = A^\dagger A = I$.
- A es normal si $AA^\dagger = A^\dagger A$.

Analogamente se define para matrices, $M \in \mathbb{C}^{n\times n}$:

- M es hermítico si M=$M^\dagger$.
- M es unitaria si $MM^\dagger = M^\dagger M = I$.
- M es normal si $MM^\dagger = M^\dagger M$.

Observación 2.1. Si A es hermítco o unitaria entonces A es nomal.

Estamos en condiciones de enunciar uno de los teoremas más importantes para operadores lineales en dimensión finita.

Teorema 2.1 (Teorema espectral)**.** *Sea A un operador hermítico, entonces A es diagonalizable.*

Corolario 2.1. *Sea A un operador normal, entonces A es diagonalizable.*

Teorema 2.2. *Sea $M \in \mathbb{C}^{n\times n}$. Equivalen:*

1. *M es unitaria.*
2. *$M^\dagger$ es unitaria.*
3. *Las columnas de M forman una base ortonormal de $\mathbb{C}^n$.*
4. *Las filas de M forman una base ortonormal de $\mathbb{C}^n$.*
5. *M representa una isometría respecto al producto escalar, es decir, $< Mx|My >=< x, y >$ para todo $x, y \in \mathbb{C}^n$.*

Corolario 2.2. *Si $\hat{A}$ y $\hat{B}$ son unitarias entonces $\hat{A} \otimes \hat{B}$ es unitaria.*

Definición 2.9. Sea $M \in \mathbb{C}^{n\times m}$. Definimos la norma inducida como

$$\|M\| = \sup_{\|x\|=1, \|x\| \in \mathbb{C}^m} (\|Mx\|).$$

Análogamente, sea $A : \mathbb{C}^n \to \mathbb{C}^m$ un operador lineal. Definimos la norma inducida como

$$\|A\| = \sup_{\|x\|=1, \|x\| \in \mathbb{C}^n} (\|A(x)\|).$$

Donde las normas de $\mathbb{C}^n, \mathbb{C}^m$ son las usuales.

Observación 2.2. Como $\{x \in \mathbb{C}^k : \|x\| = 1\}$ es compacto y las aplicaciones lineales con dominio un espacio vectorial de dimensión finita son continuas, los supremos anteriores son de hecho máximos.

Observación 2.3. Si λ es una valor propio de A y x un vector propio no nulo asociado a λ

$$|\lambda| = \|\lambda \frac{x}{\|x\|}\| = \|A(\frac{x}{\|x\|})\| \leq \|A\|$$

Los valores propios están acotados por la norma de A.

La siguiente proposición relaciona la norma de un operador con su representación matricial.

Proposición 2.3. *Sea $A : \mathbb{C}^n \to \mathbb{C}^n$ un operador lineal y B una base ortonormal de $\mathbb{C}^n$. Entonces*

$$\|A\| = \|M(A, B)\|.$$

Demostración. Si $B = B_u$, donde B_u es la base usual de $\mathbb{C}^n$, la igualdad es cierta porque para todo $x \in \mathbb{C}^n$ se verifica $A(x) = M(A, B)x$. Recordemos el significado de $M(A, B)$, dado las coordenadas de un vector según la base B, obtengo la imagen de $A(x)$ en coordenadas de B. Como B es la usual las coordenadas coinciden con el propio vector.

Supongamos que B no es la usual, sean la matriz de cambio de base $M(I, B_u \leftarrow B)$. Como B, B_u son ortonomales entonces $M(I, B_u \leftarrow B)$ es una matriz unitaria. Por la proposición 2.2 obtenemos

$$\begin{aligned}
\|M(A,B)x\| &= \sqrt{< M(A,B)x | M(A,B)x >} \\
&= \sqrt{< M(I,B,B_u)^{-1}M(A,B_u)M(I,B_u,B)x | M(I,B,B_u)^{-1}M(A,B_u)M(I,B_u,B)x >} \\
&= \sqrt{< M(I,B \leftarrow B_u)^{-1}M(A,B_u)x | M(I,B \leftarrow B_u)^{-1}M(A,B_u)x >} \\
&= \sqrt{< M(A,B_u)x | M(A,B_u)x >} \\
&= \|M(A,B_u)x\|.
\end{aligned}$$

Para no tener problemas con los márgenes se ha cambiado la notación $M(I, B \leftarrow B_u)$ por $M(I, B, B_u)$ en la anterior ecuación.

En el caso anterior vimos que $\|A\| = \|M(A, B_u)x\|$, entonces $\|A\| = \|M(A, B)x\|$. □

Supongamos f una función holomorfa en un entorno de 0, en otras palabras existe un $R > 0$ tal que $f : D(0, R) \to \mathbb{C}$, $D(0, R) = \{z \in \mathbb{C} : |z| < R\}$ y f es derivable en el sentido complejo.

Como f es holomorfa, f admite un desarrollo en serie de potencias

$$f(z) = \sum_{i=0}^{\infty} a_i z^i \ , \ \forall z \in D(0, R).$$

Dicha serie converge absolutamente en $D(0, R)$, es decir,

$$\sum_{i=0}^{\infty} |a_i||z|^i < \infty \ , \ \forall z \in D(0, R).$$

Por otro lado, dados $M \in \mathbb{C}^{k\times k}$ y $n \in \mathbb{N}$ tiene sentido calcular

$$\sum_{i=0}^{n} a_i M^i. \tag{2.1}$$

Si en $\mathbb{C}^{k\times k}$ consideramos la norma definida en 2.9, dicha norma es matricial y verifica

$$\|AB\| \leq \|A\|\|B\| \ , \ \forall A, B \in \mathbb{C}^{k\times k}. \tag{2.2}$$

Si unimos (2.1) y (2.2) nos queda

$$\|\sum_{i=0}^{n} a_i M^i\| \leq \sum_{i=0}^{n} |a_i|\|M\|^i,$$

imponiendo $\|M\| < R$

$$\|\sum_{i=0}^{n} a_i M^i\| \leq \sum_{i=0}^{n} |a_i| \|M\|^i \leq \sum_{i=0}^{n} |a_i| |R|^i \leq \sum_{i=0}^{\infty} |a_i| |z|^i < \infty.$$

Como la serie $\sum_{i=0}^{n} |a_i| \|M\|^i$ es de términos positivos y acotada, converge. Consecuentemente $\sum_{i=n}^{\infty} |a_i| \|M\|^i$ tiende a 0 si n tiende a infinito.

Sean $p < q$, $p, q \in \mathbb{N}$

$$\|\sum_{i=0}^{q} a_i M^i - \sum_{i=0}^{q} a_i M^i\| = \|\sum_{i=q+1}^{p} a_i M^i\| \leq \sum_{i=q+1}^{p} |a_i| \|M\|^i \underset{p\to\infty}{\longrightarrow} 0$$

Por tanto la serie (2.1) es de Cauchy, como $\mathbb{C}^{k\times k}$ es completo con cualquier norma, resulta que (2.1) converge cuando n tiende a infinito.

Definición 2.10. Sea f holomorfa en $D(0, R)$ con $R > 0$, $f(z) = \sum_{i=0}^{\infty} a_i z^n$,sea $M \in \mathbb{C}^{k\times k}$ tal que $\|M\| < R$. Definimos

$$f(M) = \sum_{i=0}^{\infty} a_i M^i.$$

Definición 2.11. Sea f holomorfa en $D(0, R)$ con $R > 0$, $f(z) = \sum_{i=0}^{\infty} a_i z^n$, A un operador lineal de $\mathbb{C}^k$ en $\mathbb{C}^k$, B una base ortonormal de $\mathbb{C}^k$ y $\|A\| < R$. Definimos

$$f_B(A) = \sum_{i=0}^{\infty} a_i A^i = \sum_{i=0}^{\infty} a_i M(A, B)^i.$$

Como $f_B(A)$ es una matriz, podemos asociarle una aplicación lineal donde su represenación matricial es con la base B. Provisionalmente denominaremos a dicha aplicación lineal como $f^B(A)$.

Observación 2.4. Si A es un operador lineal, $M(f^B(A), B) = f_B(A)$

Observación 2.5. Por la proposición 2.3 $f_B(A)$ y $f^B(A)$ están bien definidas para cualquier base ortonormal B.

Proposición 2.4. *Sea A una aplicación lineal, B_1 y B_2 dos bases ortonormales de $\mathbb{C}^k$. Entonces $f^{B_1}(A) = f^{B_2}(A)$.*

Demostración. Cambiando de bases,

$$\begin{aligned} M(A, B_1) &= M(I, B_1 \leftarrow B_2) M(A, B_2) M(I, B_2 \leftarrow B_1) \\ &= M(I, B_2 \leftarrow B_1)^{-1} M(A, B_2) M(I, B_2 \leftarrow B_1) \\ M(A, B_1)^i &= (M(I, B_2 \leftarrow B_1)^{-1} M(A, B_2) M(I, B_2 \leftarrow B_1))^i \\ &= M(I, B_2 \leftarrow B_1)^{-1} M(A, B_2)^i M(I, B_2 \leftarrow B_1). \end{aligned}$$

Luego podemos escribir las sumas parciales como

$$\sum_{i=0}^{n} a_i M(A, B_1)^i = M(I, B_2 \leftarrow B_1)^{-1} (\sum_{i=0}^{n} a_i M(A, B_2)^i) M(I, B_2 \leftarrow B_1).$$

Tomamos limite cuando k tiende a infito

$$f_{B_1}(A) = M(I, B_2 \leftarrow B_1)^{-1} f_{B_2}(A) M(I, B_2 \leftarrow B_1).$$

Teniendo en cuenta la observación 2.4

$$M(f^{B_1}(A), B_1) = M(I, B_2 \leftarrow B_1)^{-1} M(f^{B_2}(A), B_2)(A) M(I, B_2 \leftarrow B_1) = M(f^{B_2}(A), B_1),$$

luego tenemos dos aplicaciones lineales que sobre la misma base tienen la misma expresión matrical, por tanto $f^{B_1}(A) = f^{B_2}(A)$.

□

Observación 2.6. En base a la proposición anterior todas las aplicaciones lineales $f^B(A)$ son la misma, por tanto podemos denominar a dicha aplicación como $f(A)$.

Proposición 2.5. *Sea f holomorfa en $D(0, R)$, A un operador lineal diagonalizable con valores propios $\lambda_1, \ldots, \lambda_n$. Supongamos que $f(A)$ está bien definida, entonces*

$$M(f(A), B) = \begin{pmatrix} f(\lambda_1) & 0 & \ldots & 0 \\ 0 & f(\lambda_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & f(\lambda_n) \end{pmatrix}.$$

Donde B es una matriz ortonormal que diagonaliza a A.

Demostración. Sea B una base ortonormal tal que $M(A, B)$ es diagonal. Por definición:

$$f_B(A) = M(f(A), B) = \sum_{i=0}^{\infty} a_i M(A, B)^i = \sum_{i=0}^{\infty} a_i \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}^i = \sum_{i=0}^{\infty} a_i \begin{pmatrix} \lambda_1^i & 0 & \ldots & 0 \\ 0 & \lambda_2^i & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n^i \end{pmatrix}.$$

Desarrollando las sumas parciales

$$\sum_{i=0}^{k} a_i M(A, B)^i = \sum_{j=1}^{n} \sum_{i=0}^{k} a_i \lambda_j^i \begin{pmatrix} 0 & \ldots & \ldots & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & 0 \end{pmatrix} = \sum_{j=1}^{n} \begin{pmatrix} 0 & \ldots & \ldots & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & \ldots & \ldots & 0 \end{pmatrix} \sum_{i=0}^{k} a_i \lambda_j^i,$$

donde el 1 está en la columna y fila j-esimas.

Por otro lado, la obsevacion 2.3 nos dice que $|\lambda_j| \leq \|A\| < R$ y como f es holomorfa en un entorno de $D(0,R)$,

$$f(\lambda_j) = \sum_{i=0}^{\infty} a_i \lambda_j^i \ , \ 1 \leq j \leq n.$$

Si tendemos k a infinito:

$$M(f(A),B) = \sum_{j=1}^{n} \begin{pmatrix} 0 & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{pmatrix} f(\lambda_j) = \begin{pmatrix} f(\lambda_1) & 0 & \dots & 0 \\ 0 & f(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(\lambda_n) \end{pmatrix}.$$

□

Corolario 2.3. *Si f es holomorfa en $D(0,R)$ y A es un operador hermitiano, entonces $f(A)$ es diagonalizable.*

Definición 2.12. Dado un operador lineal $A : \mathbb{C}^n \to \mathbb{C}^n$ y un vector $V \in \mathbb{C}^n$. Definimos el valor medio esperado de A respecto al vector v como

$$\langle A \rangle = < v | A(v) > .$$

Además se define la desviación típica o incertidumbre de A como

$$\Delta A = \sqrt{\langle A^2 \rangle - \langle A \rangle^2},$$

donde $A^2 = A \circ A$.

Concretando un poco, sea $|\psi\rangle = c_1 |u_1\rangle + \cdots + c_n |u_n\rangle$ y un operador lineal A,

$$\begin{aligned} \langle A \rangle &= < |\psi\rangle \, |A(|\psi\rangle) > = \langle \psi | \, A \, |\psi\rangle = (\sum_{i=1}^{n} \overline{c_i} \, \langle u_i |) A (\sum_{i=1}^{n} c_i \, |u_i\rangle) \\ &= \sum_{i,j=1}^{n} \overline{c_i} c_j \, \langle u_i | \, A \, |u_j\rangle \, . \end{aligned}$$

.

2.2. Operador de densidad

Supongamos que tenemos un conjunto de posibles estados $|\psi_1\rangle, \dots, |\psi_n\rangle$. Si ahora elegimos al azar uno de estos elementos y medimos, queremos saber la probabilidad de medir $|0\rangle$ y la de medir $|1\rangle$. Para esto usamos el denominado operador de densidad.

2.2.1. Estados puros

Supongamos para empezar que el conjunto solo tiene un posible estado, en este caso decimos que estamos ante un estado puro. Dada una base ortonormal, supongamos que el único

elemento $|\psi\rangle$ se escribe como

$$|\psi\rangle = c_1 |u_1\rangle + \cdots + c_n |u_n\rangle .$$

Entonces la probabilidad de medir $|u_i\rangle$ es $|c_i|^2$.

Definición 2.13. Definimos el operador densidad, ρ, de $|\psi\rangle$ como:

$$\rho = |\psi\rangle \langle\psi| .$$

Observación 2.7. El operador de densidad con estado puros es una aplicación lineal.

Nos vendrá bien tener en cuenta el siguiente resultado:

Proposición 2.6 (Relación de Clausura)**.** *Sea* $B = \{|u_1\rangle , \ldots , |u_n\rangle\}$ *una base ortonormal entonces:*

$$\sum_{i=1}^{n} |u_i\rangle \langle u_i| = I.$$

Demostración. Si lo vemos como un problema de aplicaciones lineales, donde A es la aplicación lineal definida como

$$A = \sum_{i=1}^{n} |u_i\rangle \langle u_i| ,$$

vamos a ver que $M(A, B) = M(I, B)$.

$$< |u_i\rangle , A(|u_i\rangle) >= \langle u_i| A |u_i\rangle = \sum_{j=1}^{n} \langle u_j| |u_j\rangle \langle u_j| |u_i\rangle \overset{\text{Ortogonalidad}}{=} \langle u_j| |u_i\rangle .$$

Por tanto

$$M(A, B) = I = M(I, B)$$

así que $A = I$, tomando la base usual

$$M(A, B_u) = \sum_{i=1}^{n} |u_i\rangle \langle u_i| = I.$$

□

Teniendo en cuenta la proposición anterior y la definición del operador de densidad veamos otra forma de hallar el valor medio de un operador A sobre $|\psi\rangle$:

$$\begin{aligned}
\langle A\rangle &= \sum_{i,j=1}^{n} c_j \overline{c_i} \langle u_i| A |u_j\rangle = \sum_{i,j=1}^{n} \langle u_j| |\psi\rangle \langle\psi| |u_i\rangle \langle u_i| A |u_j\rangle \\
&= \sum_{i,j=1}^{n} \langle u_j| \rho |u_i\rangle \langle u_i| A |u_j\rangle = \sum_{j=1}^{n} \langle u_j| \rho (\sum_{i=1}^{n} |u_i\rangle \langle u_i|) A |u_j\rangle \\
&= \sum_{j=1}^{n} \langle u_j| \rho A |u_j\rangle = \mathrm{tr}(\rho A).
\end{aligned}$$

Veamos algunas propiedades del operador de densidad para estados puros.

Proposición 2.7. *Sea* $|\psi\rangle = c_1 |u_1\rangle + \cdots + c_n |u_n\rangle$ *y el operador de densidad* $\rho = |\psi\rangle \langle\psi|$. *Entonces:*

1. ρ *es hermitiano.*
2. $< x|\rho(x) > \geq 0$ *para todo* $x \in \mathbb{C}^n$.
3. $tr(\rho) = 1$.
4. $tr(\rho^2) = 1$.

Demostración. 1. $\rho^\dagger = (|\psi\rangle \langle\psi|)^\dagger = |\psi\rangle \langle\psi|$.

2. $< x|\rho(x) >= \langle x| |\psi\rangle \langle\psi| |x\rangle = \overline{\langle\psi| |x\rangle} \langle\psi| |x\rangle = |\langle x| |\psi\rangle|^2 \geq 0$.

3. $\mathrm{tr}(\rho) = \sum_{i=1}^{n} \langle u_i| \rho |u_i\rangle = \sum_{i=1}^{n} \langle u_i| |\psi\rangle \langle\psi| |u_i\rangle = \sum_{i=1}^{n} |c_i|^2 = 1$.

4. $\mathrm{tr}(\rho^2) = \sum_{i=1}^{n} \langle u_i| \rho^2 |u_i\rangle = \sum_{i=1}^{n} \langle u_i| |\psi\rangle \langle\psi| |\psi\rangle \langle\psi| |u_i\rangle = \sum_{i=1}^{n} \langle u_i| |\psi\rangle \langle\psi| |u_i\rangle = \sum_{i=1}^{n} |c_i|^2 = 1$.

□

2.2.2. Estados mixtos

Ahora nuestro conjunto de estados tiene n estados preparados $|\psi_1\rangle, \ldots, |\psi_n\rangle$ con una probabilidad de $p_1, \ldots, p_n$ de elegir cada uno respectivamente. Supongamos que $\sum_{i=1}^{n} p_i = 1$. Si hay al menos dos probabilidades positivas decimos que estamos ante un estado mixto, ya que en otro caso, a nivel práctico, seguimos teniendo solo un posible estado.

Definición 2.14. Sea un conjunto de estados $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$ y un conjunto de probabilidades $\{p_1, \ldots, p_n\}$. Definimos el operador de densidad para estados mixtos como

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle \langle\psi_i| .$$

Al igual que para estados puros ρ es una operación lineal.

Observación 2.8. El operador de densidad para estados mixtos es una combinación convexa de operadores de densidad para estados puros.

Obsérvese que se conservan la mayoría de las propiedades del operador de densidad para estados puros.

Proposición 2.8. *Sea* $\{|\psi_i\rangle = c_1^i |u_1\rangle + \cdots + c_n^i |u_n\rangle : 1 \leq i \leq k\}$, *un conjunto de probabilidades* $\{p_1, \ldots, p_n\}$ *y el operador de densidad* ρ *asociado. Entonces:*

1. ρ *es hermitiano.*
2. $< x|\rho(x) > \geq 0$ *para todo* $x \in \mathbb{C}^n$.
3. $tr(\rho) = 1$.

Demostración. Totalmente análoga a las propiedades del operador de densidad para estados puros, teniendo en cuenta la linealidad de la traza, producto escalar y operador adjunto.

1. $\rho^\dagger = (\sum_{i=1}^{k} p_i |\psi_i\rangle \langle\psi_i|)^\dagger = \sum_{i=1}^{k} p_i(|\psi_i\rangle \langle\psi_i|)^\dagger = \sum_{i=1}^{k} p_i |\psi_i\rangle \langle\psi_i| = \rho.$

2. $< x|\rho(x) >= \langle x| (\sum_{i=1}^{k} p_i |\psi_i\rangle \langle\psi_i|) |x\rangle = \sum_{i=1}^{k} p_i |\langle x| |\psi_i\rangle|^2 \geq 0.$

3. $\mathrm{tr}(\rho) = \sum_{i=1}^{n} \langle u_i| (\sum_{j=1}^{k} p_j |\psi_j\rangle \langle\psi_j|) |u_i\rangle = \sum_{j=1}^{k} p_j \sum_{i_1}^{n} \langle u_i| |\psi_j\rangle \langle\psi_j| |u_i\rangle = \sum_{j=1}^{k} p_j = 1.$

□

En estados mixtos no se cumple $\mathrm{tr}(\rho^2) = 1$, veamoslo por ejemplo para los estados $\{|+\rangle, |-i\rangle\}$ que definimos en (1.1) con probabilidad $\frac{1}{4}$ y $\frac{3}{4}$ respectivamente. Primero calculemos el operador de densidad

$$\begin{aligned}\rho &= \frac{1}{4}|+\rangle\langle+| + \frac{3}{4}|-i\rangle\langle-i| = \frac{1}{4}(\frac{|0\rangle+|1\rangle}{\sqrt{2}})(\frac{\langle 0|+\langle 1|}{\sqrt{2}}) + \frac{3}{4}(\frac{|0\rangle - i|1\rangle}{\sqrt{2}})(\frac{\langle 0|+i\langle 1|}{\sqrt{2}}) \\ &= \frac{1}{2}(\frac{1}{4}(|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|) + \frac{3}{4}(|0\rangle\langle 0| + i|0\rangle\langle 1| - i|1\rangle\langle 0| + |1\rangle\langle 1|)) \\ &= \frac{1}{2}(\frac{4|0\rangle\langle 0| + (1+3i)|0\rangle\langle 1| + (1-3i)|1\rangle\langle 0| + 4|1\rangle\langle 1|}{4}),\end{aligned}$$

consecuentemente ρ^2

$$\rho^2 = \frac{26|0\rangle\langle 0| + (8+24i)|0\rangle\langle 1| + (8-24i)|1\rangle\langle 0| + 26|1\rangle\langle 1|}{64}.$$

En este caso $\mathrm{tr}(\rho^2) = \frac{52}{64} < 1$.
Este hecho nos hace sospechar de la siguiente caracterización de estados puros y mixtos.

Proposición 2.9. *Sea ρ un operador de densidad. Entonces:*

- $tr(\rho^2) \leq 1$.
- $tr(\rho^2) = 1 \Leftrightarrow \rho$ *proviene de estado puro.*

Demostración. Veamos el primer punto:

- Si ρ proviene de un estado puro el resultado es cierto, por tanto, supongamos que ρ proviene de un estado mixto. Sea $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$ con probabilidades $\{p_1, \ldots, p_n\}$ tales que

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle \langle\psi_i|.$$

  Entonces

$$\rho^2 = (\sum_{i=1}^{n} p_i |\psi_i\rangle \langle\psi_i|)^2 = \sum_{i,j=1}^{n} p_i p_j |\psi_i\rangle \langle\psi_i| |\psi_j\rangle \langle\psi_j|.$$

Si calculamos ahora la traza de ρ^2

$$\begin{aligned}\mathrm{tr}(\rho^2) &= \mathrm{tr}(\sum_{i,j=1}^{n} p_i p_j \left|\psi_i\right\rangle \left\langle\psi_i\right| \left|\psi_j\right\rangle \left\langle\psi_j\right|) = \sum_{i,j=1}^{n} p_i p_j \mathrm{tr}(\left|\psi_i\right\rangle \left\langle\psi_i\right| \left|\psi_j\right\rangle \left\langle\psi_j\right|) \\ &= \sum_{i,j=1}^{n} p_i p_j \left\langle\psi_j\right| \left|\psi_i\right\rangle \left\langle\psi_i\right| \left|\psi_j\right\rangle = \sum_{i,j=1}^{n} p_i p_j \left\langle\psi_j\right| \left|\psi_i\right\rangle \overline{\left\langle\psi_j\right| \left|\psi_i\right\rangle} \\ &= \sum_{i,j=1}^{n} p_i p_j |\left\langle\psi_j\right| \left|\psi_i\right\rangle|^2 \leq \sum_{i,j=1}^{n} p_i p_j |\left\langle\psi_i\right| |^2 |\left|\psi_j\right\rangle|^2 \\ &= \sum_{i,j=1}^{n} p_i p_j = (\sum_{i=1}^{n} p_i)^2 = 1.\end{aligned}$$

En la anterior igualdad hemos usado la propiedad de que si A es un operador lineal entonces $\mathrm{tr}(A \left|u\right\rangle \left\langle v\right|) = \left\langle u\right| A \left|v\right\rangle$, tomando $A = \left|\psi_i\right\rangle \left\langle\psi_i\right|$.

- Ahora veamos $\mathrm{tr}(\rho^2) = 1 \Leftrightarrow \rho$ proviene de estado puro. La implicación de izquierda a derecha se sigue de en la proposición 2.7, así que demostremos la otra implicación.

  Como ρ es hermitiano es diagonalizable por el teorema espectral. Además, como ρ es semidefinido positivo, sus valores propios son reales y mayores o iguales que 0, como $\mathrm{tr}(\rho) = 1$ los valores propios están acotados por 1. En resumen, para cada valor propio λ_i es verifica $0 \leq \lambda_i \leq 1$ con $1 \leq i \leq n$. Sea una base ortonormal que diagonaliza ρ, $B = \{\left|u_1\right\rangle, \ldots, \left|u_n\right\rangle\}$, donde $\left|u_i\right\rangle$ es el vector propio de λ_i. Definimos:

$$\left|\psi\right\rangle = \sqrt{\lambda_1} \left|u_1\right\rangle + \cdots + \sqrt{\lambda_n} \left|u_n\right\rangle$$

$$\left|\psi\right\rangle \left\langle\psi\right| = (\sqrt{\lambda_1} \left|u_1\right\rangle + \cdots + \sqrt{\lambda_n} \left|u_n\right\rangle)(\sqrt{\lambda_1} \left\langle u_1\right| + \cdots + \sqrt{\lambda_n} \left\langle u_n\right|) = \sum_{i=1}^{n} \lambda_i \left|u_i\right\rangle \left\langle u_i\right|.$$

  Si consideramos la aplicación definida por $\left|\psi\right\rangle \left\langle\psi\right|$

$$\left\langle u_j\right| \left|\psi\right\rangle \left\langle\psi\right| \left|u_i\right\rangle = \lambda_i \left\langle u_j\right| \left|u_i\right\rangle,$$

  por tanto, la expresión matricial de $\left|\psi\right\rangle \left\langle\psi\right|$ en la base B es

$$M(\left|\psi\right\rangle \left\langle\psi\right|, B) = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} = M(\rho, B).$$

  Como la expresión matrical de ρ y $\left|\psi\right\rangle \left\langle\psi\right|$ en la base B coinciden, entonces como aplicaciones lineales son la misma. Luego podemos escribir $\rho = \left|\psi\right\rangle \left\langle\psi\right|$ y concluimos que ρ proviene de un estado puro.

□

Observación 2.9. Se verifica la siguiente caracterización para estados puros y mixtos a partir de ρ.

- $\text{tr}(\rho^2) = 1 \Leftrightarrow$ estados puro.
- $\text{tr}(\rho^2) < 1 \Leftrightarrow$ estado mixto.

Al inicio de la sección comentábamos que con el operador de densidad podemos calcular la probabilidad de medir $|0\rangle$ y $|1\rangle$, la argumentación formal la daremos en la siguiente sección, pero de momento indicamos que en la base usual ρ se puede escribir como:

$$\rho = \begin{pmatrix} \langle 0|\,\rho\,|0\rangle & \langle 0|\,\rho\,|1\rangle \\ \langle 1|\,\rho\,|0\rangle & \langle 1|\,\rho\,|1\rangle \end{pmatrix}.$$

La probabilidad de medir $|0\rangle$ es $\langle 0|\,\rho\,|0\rangle$ y la de medir $|1\rangle$ es $\langle 1|\,\rho\,|1\rangle$.

2.2.3. Operador de densidad reducido

Hasta ahora hemos trabajado en un $\mathbb{C}^n$ arbitrario. Para dar sentido al operador de densidad reducido suponemos que estamos en el producto tensorial, es decir, trabajamos en, $\mathbb{C}_1^n \otimes \mathbb{C}_2^m$. Si tenemos un conjunto de probabilidades $\{p_1, \ldots, p_h\}$, sabemos que el operador de densidad se define como:

$$\rho = \sum_{i=1}^{h} p_i\,|\psi_i\rangle\,\langle\psi_i|$$

Definición 2.15. Dado un operador de densidad ρ en $\mathbb{C}_1^n \otimes \mathbb{C}_2^m$ y su expresión matricial en una base

$$\rho = \begin{pmatrix} a_{11,11} & a_{11,12} & \ldots & a_{11,1m} & a_{11,21} & \ldots & a_{11,nm} \\ a_{12,11} & a_{12,12} & \ldots & a_{12,1m} & a_{12,21} & \ldots & a_{12,nm} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{21,11} & a_{21,12} & \ldots & a_{21,1m} & a_{21,21} & \ldots & a_{21,nm} \\ a_{22,11} & a_{22,12} & \ldots & a_{22,1m} & a_{22,21} & \ldots & a_{22,nm} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{nm,11} & a_{nm,12} & \ldots & a_{nm,1m} & a_{nm,21} & \ldots & a_{nm,nm} \end{pmatrix} = \{a_{kl,ij}\},$$

definimos el operador de densidad reducido a $\mathbb{C}_1^n = H$, denotado por ρ_H, como

$$\rho_H = \begin{pmatrix} \sum_{j=1}^{m} a_{1j,1j} & \sum_{j=1}^{m} a_{1j,2j} & \ldots & \sum_{j=1}^{m} a_{1j,nj} \\ \sum_{j=1}^{m} a_{2j,1j} & \sum_{j=1}^{m} a_{2j,2j} & \ldots & \sum_{j=1}^{m} a_{2j,nj} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{j=1}^{m} a_{nj,1j} & \sum_{j=1}^{m} a_{nj,2j} & \ldots & \sum_{j=1}^{m} a_{nj,nj} \end{pmatrix}.$$

Observación 2.10. Dadas $\{u_1 \ldots, u_n\}$ y $\{v_1, \ldots, v_m\}$ bases ortonormales de $\mathbb{C}_1^n$ y $\mathbb{C}_2^m$, entonces

$$\begin{aligned} a_{kl,ij} &= \langle u_k \otimes v_l|\,\rho\,|u_i \otimes v_j\rangle\,, \\ \sum_{j=1}^{m} a_{kj,lj} &= \sum_{j=1}^{m} \langle u_k \otimes v_j|\,\rho\,|u_i \otimes v_j\rangle = \sum_{j=1}^{m} \langle v_j|\,\langle u_k|\,\rho\,|u_i\rangle\,|v_j\rangle \end{aligned}$$

y el operador de densidad reducido se puede expresar como

$$\rho_H = \sum_{j=1}^{m} \langle v_j | \rho | v_j \rangle .$$

Supongamos, por ejemplo, que estamos trabajando en $\mathbb{C}^2 \otimes \mathbb{C}^2$ con la base ortonormal $\{|0\rangle, |1\rangle\}$. Indicamos con 1 y 2 si estamos en el primer espacio del producto tensorial o en el segundo. Tomemos el siguiente qubit que ya vimos en (1.4)

$$|\psi\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}} = \frac{|0^1\rangle |1^2\rangle + |1^1\rangle |0^2\rangle}{\sqrt{2}}.$$

El operador de densidad es:

$$\begin{aligned}\rho = |\psi\rangle \langle\psi| &= \left(\frac{|0^1\rangle |1^2\rangle + |1^1\rangle |0^2\rangle}{\sqrt{2}}\right)\left(\frac{\langle 0^1| \langle 1^2| + \langle 1^1| \langle 0^2|}{\sqrt{2}}\right) \\ &= \frac{|0^1\rangle |1^2\rangle \langle 0^1| \langle 1^2| + |0^1\rangle |1^2\rangle \langle 1^1| \langle 0^2| + |1^1\rangle |0^2\rangle \langle 0^1| \langle 1^2| + |1^1\rangle |0^2\rangle \langle 1^1| \langle 0^2|}{2}.\end{aligned}$$

Si lo reducimos al espacio 1, obtenemos:

$$\begin{aligned}\rho_H &= \langle 0^2| \rho |0^2\rangle + \langle 1^2| \rho |1^2\rangle , \\ \langle 0^2| \rho |0^2\rangle &= \frac{\langle 0^2| |0^1\rangle |1^2\rangle \langle 0^1| \langle 1^2| |0^2\rangle + \langle 0^2| |0^1\rangle |1^2\rangle \langle 1^1| \langle 0^2| |0^2\rangle}{2} \\ &\quad + \frac{\langle 0^2| |1^1\rangle |0^2\rangle \langle 0^1| \langle 1^2| |0^2\rangle + \langle 0^2| |1^1\rangle |0^2\rangle \langle 1^1| \langle 0^2| |0^2\rangle}{2} \\ &= \frac{\langle 0^2| |1^2\rangle |0^1\rangle \langle 0^1| \langle 1^2| |0^2\rangle + \langle 0^2| |1^2\rangle |0^1\rangle \langle 1^1| \langle 0^2| |0^2\rangle}{2} \\ &\quad + \frac{\langle 0^2| |0^2\rangle |1^1\rangle \langle 0^1| \langle 1^2| |0^2\rangle + \langle 0^2| |0^2\rangle |1^1\rangle \langle 1^1| \langle 0^2| |0^2\rangle}{2} \\ &= \frac{|1^1\rangle \langle 1^1|}{2}, \\ \langle 1^2| \rho |1^2\rangle &= \frac{|0^1\rangle \langle 0^1|}{2}, \\ \rho_H &= \frac{|0^1\rangle \langle 0^1| + |1^1\rangle \langle 1^1|}{2}.\end{aligned}$$

2.3. Medir qubits

Cada vez que tenemos un qubit en el estado $|\psi\rangle = a|0\rangle + b|1\rangle$, sabemos que la probabilidad de medir $|0\rangle$ es $|a|^2$ y de la de medir $|1\rangle$ es $|b|^2$. Además, al medir forzamos al qubit a tomar el estado $|0\rangle$ o $|1\rangle$, y si no conocemos la forma concreta cuando lo medimos, no podemos averiguar los valores de a y b.

Cuando medimos, lo hacemos desde una base ortonormal $B = \{|u_1\rangle, \ldots, |u_n\rangle\}$ y tras la

medición obtenemos un nuevo qubit que está en combinación lineal de $B' \subsetneq B$. Para describir una medición vamos a usar las proyecciones lineales.

Antes de poder definir una proyección lineal, hemos de justificar que dado un espacio de Hilbert de dimensión finita, dim(H)=n, un subespacio W suyo y $v \in H$, entonces, existen unos únicos $w \in W$ y $s \in W^{\perp} = \{x \in H :< x, y >= 0 \; \forall y \in W\}$ verificando $v = w + s$.

Supongamos que W es un subespacio propio, sea $B_W = \{w_1, \ldots, w_m\}$ una base ortogonal de W, $m < n$. Completamos B_W hasta obtener una base ortogonal de H, $B = \{w_1, \ldots, w_m, v_1, \ldots, v_{n-m}\}$. Como $v_1, \ldots, v_{n-m}$ son ortogonales a B_W, entonces son ortogonales a W y por definición perteneceran a $W^{\perp}$.

Dado $v \in H$, como B es una base existen unos únicos coeficientes $a_1, \ldots, a_n \in K$ verificando

$$v = a_1 w_1 + \cdots + a_m w_m + a_{m+1} v_1 + \cdots + a_n v_{n-m}$$

Definiendo $w = a_1 w_1 + \cdots + a_m w_m$ perteneciente a W y $s = a_{m+1} v_1 + \cdots + a_n v_{n-m}$ pertenciente a $W^{\perp}$, se obtiene que $v = w + s$. Como los coeficientes $a_1, \ldots, a_n \in K$ son únicos, entonces w y s son únicos. En caso de que W sea H o $\{0\}$, $W^{\perp}$ es $\{0\}$ o H y lo anterior sigue siendo cierto.

Definición 2.16. Sea un espacio de Hilbert H de dimensión finita y un subespacio suyo W. Dado $v \in H$ unos únicos $w \in W$ y $s \in W^{\perp}$ tales que $v = w + s$. Definimos la proyección de H en W como:

$$P(v) = P(w + s) = w.$$

Las proyecciones son aplicaciones hermíticas.

Observación 2.11. De la definición es inmediato que $P^2 = P$.

La definición anterior es aplicable para espacios de Hilbert arbitrarios si W es un subespacio cerrado. Como nos vamos a restringir a espacios de dimensión finita, concretamente C^n con $n \in \mathbb{N}$, no probaremos que la definición tiene sentido en espacios de Hilbert arbitrarios con W subespacio cerrado.

Proposición 2.10. *Dado $|u\rangle \in \mathbb{C}^{n+1}$ vector unitario, sea el subespacio generado por $|u\rangle$, $L(|u\rangle) = \{\lambda |u\rangle : \lambda \in \mathbb{C}\} = W$. Entonces se verifica:*

$$P_W = |u\rangle \langle u|$$

Demostración. Partiendo de $|u\rangle$ completemos hasta tener una base ortonormal $B = \{|u\rangle, |v_1\rangle, \ldots, |v_n\rangle\}$, como la imagen por la base B de ambas aplicaciones lineales son iguales:

$$\begin{aligned} P_W(|u\rangle) &= |u\rangle \\ P_W(|v_i\rangle) &= 0 \\ (|u\rangle \langle u|)(|u\rangle) &= |u\rangle \\ (|u\rangle \langle u|)(|v_i\rangle) &= 0, \end{aligned}$$

ambas aplicaciones serán iguales. □

Observación 2.12. Por tanto el operador de densidad es una proyección o combinación convexa de proyecciones.

Definición 2.17. Dos proyecciones P_1, P_2 son ortogonales si se verifica

$$(P_1 \circ P_2)(v) = 0 \ , \ \forall v \in \mathbb{C}^n.$$

Diremos que un conjunto de proyecciones no nulas $\{P_1, \ldots, P_k\}$ forman un conjunto completo de proyecciones ortogonales si además de ser ortogonales entre sí verifican

$$\sum_{i}^{k} P_i = I.$$

Nosotros medimos respecto a un conjunto completo de proyecciones ortogonales $\{P_1, \ldots, P_k\}$ el qubit ,$|\psi\rangle$, colapsa o se proyecta por alguna de las proyecciones del conjunto. La probabilidad con la que se usa la proyeccion P_i es:

$$Pr(i) = |P_i(|\psi\rangle)|^2 = (P_i(|\psi\rangle))^\dagger (P_i(|\psi\rangle)) = \langle\psi| P_i P_i |\psi\rangle = \langle\psi| P_i |\psi\rangle = \mathrm{tr}(P_i |\psi\rangle \langle\psi|) = \mathrm{tr}(P_i \rho).$$

En lo anterior estamos interpretando a P_i en su expresión matricial sobre la base usual y ρ es el operador de densidad del estado $|\psi\rangle$.

Proposición 2.11. *Dado un conjunto completo de proyecciones ortogonales $\{P_1, \ldots, P_k\}$ y un qubit $|\psi\rangle$. Entonces:*

1. $Pr(i) \leq 1$.
2. $\sum_{i=1}^{k} Pr(i) = 1$.

Demostración. 1. Cada P_j tendrá asociado un subespacio vectorial al que proyecta. De cada uno tomamos una base ortonormal $\{u_1^j, \ldots, u_{n_j}^j\}$ y la unión de todas las bases será una base ortonormal de todo el espacio. Luego podemos expresar $|\psi\rangle$ en dicha base

$$|\psi\rangle = \sum_{j=1}^{k} \sum_{h=1}^{n_j} c_{j,h} u_h^j,$$

$$1 = |\,|\psi\rangle\,|^2 = \sum_{j=1}^{k} \sum_{h=1}^{n_j} |c_{j,h}|^2.$$

Al calcular la imagen por P_i y la consecuente probabilidad, obtenemos

$$P_i |\psi\rangle = \sum_{h=1}^{n_i} c_{i,h} u_h^i,$$

$$Pr(i) = |P_i |\psi\rangle\,|^2 = \sum_{h=1}^{n_i} |c_{i,h}|^2 \leq \sum_{j=1}^{k} \sum_{h=1}^{n_j} |c_{j,h}|^2 = 1.$$

2. $\sum_{i=1}^{k} Pr(i) = \sum_{i=1}^{k} \langle\psi| P_i |\psi\rangle = \langle\psi| \sum_{i}^{k} P_i |\psi\rangle = \langle\psi|\,|\psi\rangle = 1.$

□

Consideremos un observable A, un operador hermítico por postulado 2, aplicando el teorema de diagonalización espectral existe una base ortonormal $\{|u_1\rangle, \ldots, |u_n\rangle\}$ donde $|u_i\rangle$ está asociado a un valor propio a_i. A se puede expresar como

$$A = \sum_{i=1}^{n} a_i |u_i\rangle \langle u_i| = \sum_{i=1}^{n} a_i P_i,$$

donde $P_i = |u_i\rangle \langle u_i|$. Hemos podido descomponer cualquier observable como suma de un conjunto completo de proyecciones ortogonales.

Dado $|\psi\rangle$ se verifica

$$|\psi\rangle = \sum_{i=1}^{n} c_i |u_i\rangle ,$$

llamamos a $c_i = \langle u_i| |\psi\rangle$, amplitud de probabilidad de medir a_i. La probabilidad de medir a_i es

$$Pr(a_i) = |c_i|^2 = |\langle u_i| |\psi\rangle|^2.$$

Si $|\psi\rangle$ no tenia norma uno desde el inicio, dividimos por su norma supuesto que no sea o. Por otro lado también se puede calcular el valor medio del operador A sobre el estado $|\psi\rangle$ como:

$$\langle A\rangle = \langle\psi| A |\psi\rangle = \langle\psi| (\sum_{i=1}^{n} a_i P_i) |\psi\rangle = \sum_{i=1}^{n} a_i \langle\psi| P_i |\psi\rangle = \sum_{i=1}^{n} a_i \mathrm{tr}(P_i |\psi\rangle \langle\psi|) = \sum_{i=1}^{n} a_i Pr(i).$$

Ahora sabemos con qué probabilidad vamos a medir alguno de los posibles estados, pero también queremos saber a qué estado hemos llegado. Hagamos notar que tras cada medición forzamos al qubit a proyectarse. Supongamos que se ha usado una proyección P_i con $Pr(i) > 0$ al estado $|\psi\rangle$, el nuevo estado $|\psi'\rangle$ es:

$$|\psi'\rangle = \frac{P_i |\psi\rangle}{\sqrt{\langle\psi| P_i |\psi\rangle}} = \frac{P_i |\psi\rangle}{\sqrt{Pr(i)}} = \frac{P_i |\psi\rangle}{\sqrt{|P_i(|\psi\rangle)|^2}} = \frac{P_i |\psi\rangle}{|P_i |\psi\rangle|}.$$

Lo que hacemos es aplicar la proyección P_i y luego normalizamos porque los estados de los qubits siempre tienen norma 1. Cabe discutir que si $P_i |\psi\rangle = 0$ entonces $Pr(i) = 0$ y por tanto cuando se mide el estado no se aplicaría esta proyección, así que no estamos dividiendo por 0. También podemos calcular el operador de densidad del nuevo estado a partir del operador de densidad de $|\psi\rangle$

$$\rho' = |\psi'\rangle \langle\psi'| = \frac{P_i |\psi\rangle \langle\psi| P_i}{\mathrm{tr}(P_i |\psi\rangle \langle\psi|)} = \frac{P_i \rho P_i}{\mathrm{tr}(P_i \rho)}.$$

Veamos ahora una forma de generalizar las mediciones.

Definición 2.18. Dado un operador M_m, donde m denota un posible resultado tras medir, y un estado $|\psi\rangle$, la probabilidad de usar este operador para medir es:

$$Pr(m) = \langle\psi| M^\dagger M |\psi\rangle = \mathrm{tr}(M^\dagger M |\psi\rangle \langle\psi|) = \mathrm{tr}(M^\dagger M \rho).$$

Observación 2.13. El operador $M^\dagger M$ es hermítico.

Observación 2.14. Las proyecciones se pueden ver como mediciones de este tipo.

Igual que antes podemos tener un conjunto $\{M_1, \ldots, M_k\}$ de operadores que verifican una relación de clausura, es decir

$$\sum_{m=1}^{k} M^\dagger M = I.$$

En cuyo caso, cuando medimos respecto a estos operadores siempre se aplicará uno de ellos con probabilidad $Pr(m)$. Tras medir el estado vuelve a colapsar, es decir, se aplica alguno de los operadores M_m. El nuevo estado $|\psi'\rangle$ y el consecuente operador de densidad ρ' son:

$$|\psi'\rangle = \frac{M_m |\psi\rangle}{\sqrt{\langle\psi| M_m^\dagger M_m |\psi\rangle}},$$

$$\rho' = \frac{M_m \rho M_m^\dagger}{\mathrm{tr}(M_m^\dagger M_m \rho)}.$$

2.3.1. POVM

Definición 2.19. Un POVM (positive operator-valued measurement) es un conjunto de operadores hermíticos semidefinidos positivos $\{E_1, \ldots, E_k\}$ verificando la siguiente relación de clausura

$$\sum_{m=1}^{k} E_m = I.$$

Observación 2.15. Podemos construir un POVM dando una serie $\{M_1, \ldots, M_k\}$, operadores semidefinidos positvos tomando $E_m = M_m^\dagger M_m$.

Observación 2.16. Un conjunto completo de proyecciones ortogonales es un POVM.

Hay varias ventajas que ofrece un POVM sobre un conjunto completo de proyecciones ortogonales. La primera de ellas es ser capaces de medir en bases no ortonormales. Por ejemplo tomemos $|\psi\rangle$ y $|\phi\rangle$ dos estados distintos y no ortogonales, entonces $|\langle\psi|\phi\rangle| = \cos(\theta)$, definimos:

$$E_1 = \frac{I - |\psi\rangle\langle\psi|}{1+\cos(\theta)}\ ,\ E_2 = \frac{I - |\phi\rangle\langle\phi|}{1+\cos(\theta)}\ ,\ E_3 = I - E_1 - E_2.$$

El conjunto $\{E_1, E_2, E_3\}$ es un POVM. Veamos las probabilidades asociadas a cada estado con el operador E_1

$$\langle\psi| E_1 |\psi\rangle = \frac{\langle\psi|\psi\rangle - \langle\psi|\psi\rangle\langle\psi|\psi\rangle}{1+\cos(\theta)} = \frac{1-1}{1+\cos(\theta)},$$
$$\langle\phi| E_1 |\phi\rangle = \frac{\langle\phi|\phi\rangle - \langle\phi|\psi\rangle\langle\psi|\phi\rangle}{1+\cos(\theta)} = \frac{1-|\langle\psi|\phi\rangle|^2}{1+\cos(\theta)} = \frac{1-\cos^2(\theta)}{1+\cos(\theta)} = 1-\cos(\theta).$$

Con E_1 tenemos una probabilidad de identificar al estado $|\phi\rangle$ de $1-\cos(\theta)$ y al estado $|\psi\rangle$ de 0. Análogamente ocurre con E_2 intercambiando las probabilidades. Estos operadores nos dan metodos para distinguir entre estados no ortogonales bajo una cierta imprecisión. Si al medir se aplica al estado el operador E_3, no sacamos información relevante sobre el estado.

Otra aplicación de un POVM es la medición débil: medir un estado procurando cambiarlo

lo mínimo posible a cambio de obtener menos información. Vamos a describir un POVM que realiza mediciones débiles para solo un qubit. Sean el qubit $|\psi\rangle$, E_1, E_2 y $0 \leq \epsilon \leq 1$:

$$\begin{aligned} |\psi\rangle &= a\,|0\rangle + b\,|1\rangle\,, \\ E_1 &= |0\rangle\langle 0| + (1-\epsilon)\,|1\rangle\langle 1|\,, \\ E_2 &= \epsilon\,|1\rangle\langle 1|\,. \end{aligned}$$

Fácilmente se comprueba que E_1 y E_2 verifican la relación de clausura, por linealidad ambos son hermíticos y son semidefinidos positivos porque los valores propios de E_1 son $\{1, (1-\epsilon)\}$ y los de E_2 son $\{0, 1\}$ siendo en ambos casos $|0\rangle$ y $|1\rangle$ vectores propios.

La probabilidad de obtener una medición con E_1 y con E_2 es

$$\begin{aligned} \langle\psi|\,E_1\,|\psi\rangle &= (\overline{a}\,\langle 0| + \overline{b}\,\langle 1|)\,|0\rangle\langle 0| + (1-\epsilon)\,|1\rangle\langle 1|\,(a\,|0\rangle + b\,|1\rangle) \\ &= (\overline{a}\,\langle 0| + \overline{b}(1-\epsilon)\,\langle 1|)(a\,|0\rangle + b\,|1\rangle) = |a|^2 + |b|^2(1-\epsilon), \\ \langle\psi|\,E_2\,|\psi\rangle &= (\overline{a}\,\langle 0| + \overline{b}\,\langle 1|)\epsilon\,|1\rangle\langle 1|\,(a\,|0\rangle + b\,|1\rangle) = \overline{b}\epsilon\,\langle 1|\,(a\,|0\rangle + b\,|1\rangle) \\ &= |b|^2\epsilon. \end{aligned}$$

Y los estados tras cada medición serían

$$\begin{aligned} |\psi^1\rangle &= \frac{E_1\,|\psi\rangle}{\sqrt{\langle\psi|\,E_1\,|\psi\rangle}} = \frac{a\,|0\rangle + b(1-\epsilon)\,|1\rangle}{\sqrt{|a|^2 + |b|^2(1-\epsilon)}}, \\ |\psi^2\rangle &= \frac{E_2\,|\psi\rangle}{\sqrt{\langle\psi|\,E_2\,|\psi\rangle}} = \frac{b\epsilon\,|1\rangle}{|b|\sqrt{\epsilon}}. \end{aligned}$$

Luego si ϵ es muy cercano a 0 entonces podemos medir casi sin afectar al estado $|\psi\rangle$ con mucha probabilidad, y con una probabilidad baja tras la medición el estado deja de estar en superposición y colapsa en $|1\rangle$.

2.4. Puertas lógicas

Sabemos como medir qubits con operadores y podemos describirlos con el operador de densidad, ahora vamos a introducir las operaciones que podemos hacer con ellos. Estas operaciones las llamamos puertas lógicas en un análogo a la computación clásica.

Exigimos a las puertas lógicas que sean operaciones lineales de $\mathbb{C}^n$ en $\mathbb{C}^n$ y que lleven qubits a qubits. Es decir, que lleven vectores de norma 1 a vectores de norma 1, por tanto las puertas lógicas son isometrías, equivalentemente, aplicaciones lineales unitarias.

Definición 2.20. Una puerta lógica es una aplicación lineal $A : \mathbb{C}^n \to \mathbb{C}^n$ unitaria.

Observación 2.17. Como las aplicaciones unitarias tienen inversa implica que toda puerta lógica es invertible. Es decir, sabiendo su salida podemos saber su entrada. Por tanto un circuito cuańtico es reversible mientras no se haya hecho ninguna medición.

La primera aplicación que vamos a ver de las puertas lógicas es el principio de no clonación. Cada vez que medimos perdemos el estado original, ¿por qué no copiamos dicho estado en otros qubits y luego continuamos operando con la copia pero con más información?. Si lo traducimos a algo prácticable sería: tengo dos sistemas de qubits del mismo tamaño, uno en el estado $|\psi\rangle$ y el otro en el estado $|\phi\rangle$, y quiero dejar ambos sistems en el mismo estado, por

ejemplo $|\psi\rangle$. Como estoy operando con qubits, la copia se tiene que realizar a traves de una puerta lógica, pero dicha puerta no existe.

Teorema 2.3 (Principio de no clonación)**.** *Dado dos qubits, el primero en el estado $|\psi\rangle$ y el segundo $|\phi\rangle$, con $|\psi\rangle \neq |\phi\rangle$. Entonces no existe ninguna puerta lógica que permita copiar un estado en el otro.*

Demostración. Supongamos que existe dicha puerta lógica, entonces existe una isometría lineal $A : \mathbb{C}^n \to \mathbb{C}^n$ tal que:

$$A|\psi\rangle = |\psi\rangle \ , \ A|\phi\rangle = |\psi\rangle .$$

Por linealidad

$$A(|\psi\rangle - |\phi\rangle) = A|\psi\rangle - A|\phi\rangle = |\psi\rangle - |\psi\rangle = 0,$$

luego el vector $|\psi\rangle - |\phi\rangle$ está en el núcleo, como A es una isometría deducimos $|\psi\rangle = |\phi\rangle$ en contra de nuestra hipótesis. Por tanto dicha puerta lógica no existe.

□

2.4.1. Un qubit

Empecemos por las puertas lógicas que tenemos para un qubit, las cuatro siguientes son de las más relevantes:

Definición 2.21 (Puertas de Pauli)**.** Denominamos puertas de Pauli a las puertas lógicas definidas por las matrices en la base usual:

$$\sigma_0 = \sigma_I = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \sigma_1 = \sigma_X = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\sigma_2 = \sigma_Y = Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \sigma_Z = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Observación 2.18. Las puertas de Pauli $\{\sigma_I, \sigma_X, \sigma_Y, \sigma_Z\}$ constituye una base ortonormal de las aplicaciones lineales de $\mathbb{C}^2$ en $\mathbb{C}^2$.

Del hecho que forman una base se puede obtener una curiosa aplicación junto al operador de densidad ρ. Como ρ es una aplicación hermítica se puede escribir matricialmente de la siguiente forma:

$$\rho = \begin{pmatrix} a+b & c-id \\ c+id & a-b \end{pmatrix} \ , \ \mathrm{tr}(\rho) = 2a = 1 \implies a = \frac{1}{2}.$$

Si denominamos $x = 2c$, $y = 2d$, $z = 2b$

$$\rho = \frac{1}{2}(\sigma_I + x\sigma_X + y\sigma_Y + z\sigma_Z) = \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix}.$$

Definimos el vector de Bloch como $\vec{S} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Primero veamos que la norma de este vector es siempre menor o igual que 1 usando el polinomio característico de ρ

$$\det(\rho - \lambda I) = \lambda^2 - \mathrm{tr}(\rho)\lambda + (det)(\rho) = \lambda^2 - \lambda + (1 - z^2 - x^2 - y^2) = \lambda^2 - \lambda + (1 - |\vec{S}|^2).$$

Teniendo en cuenta que ρ es hermitiana, semidefinida positiva y con traza igual a 1, todos los valores propios de ρ están ente 0 y 1. Y como es un polinomio de segundo grado con coeficiente lider positivo entonces:

$$1 - |\vec{S}|^2 \geq 0 \implies 1 \geq |\vec{S}|^2.$$

Proposición 2.12. *Sea ρ un operador de densidad en $\mathbb{C}^2$ y $\vec{S}$ su vector de Bloch asociado. Entonces ρ proviene de un estado puro si y solo si $|\vec{S}|^2 = 1$.*

Demostración. Si ρ proviene de un estado puro es porque $\rho = |\psi\rangle\langle\psi|$, como

$$\rho|\psi\rangle = |\psi\rangle\langle\psi|\psi\rangle = |\psi\rangle$$

el 1 es un valor propio de ρ. Sustituyendo en su polinomio característico, $1 - |\vec{S}|^2 = 0 \implies 1 = |\vec{S}|^2$.

Recíprocamente, si $|\vec{S}|^2 = 1$, los valores propios de ρ son 1 y 0. Tomando $\{|\psi\rangle, |\phi\rangle\}$ una base ortonormal que diagonaliza a ρ, donde $|\psi\rangle$ está asociado al valor propio 1, por coincidir en una base, $\rho = |\psi\rangle\langle\psi|$. □

Observación 2.19. De la prueba se deduce que si ρ tiene un valor propio distinto de 1 si y solo si define un estado mixto.

Como toda aplicación hermítica, las puertas de Pauli son diagonalizables. El valor propio de I es el $\{1\}$ y una base ortonormal que lo diagonaliza es $\{|0\rangle, |1\rangle\}$, los valores propios de Z son $\{1, -1\}$ y una base ortonomal asociada es $\{|0\rangle, |1\rangle\}$. Para las puertas X e Y la observación no es tan directa, recordemos los estados definidos en (1.1):

$$X|+\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = |+\rangle$$

$$X|-\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = -|-\rangle$$

$$Y|i\rangle = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ i\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ i\frac{1}{\sqrt{2}} \end{pmatrix} = |i\rangle$$

$$Y|-i\rangle = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -i\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ i\frac{1}{\sqrt{2}} \end{pmatrix} = -|-i\rangle.$$

Por tanto los valores propios de X e Y son $\{1, -1\}$ y la base que diagonaliza es $\{|+\rangle, |-\rangle\}$ en X y $\{|i\rangle, |-i\rangle\}$ en Y.

En 2.10 vimos la definición de la función de una aplicación lineal. Si tomamos como función holomorfa la exponencial $f(z) = e^{-i\theta z}$, esta función es holomorfa en todo $\mathbb{C}$, por tanto podemos calcular el exponente de cualquier aplicación lineal o matriz. Además, en 2.5 vimos una forma eficaz de calcular la función de una matriz diagonalizable. Juntando los tres hechos anteriores la exponenciación de las puertas de Pauli es:

$$\begin{aligned}
e^{-i\theta I} &= \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} = M(e^{-i\theta I}, \{|0\rangle, |1\rangle\}) \\
e^{-i\theta Z} &= \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix} = M(e^{-i\theta Z}, \{|0\rangle, |1\rangle\}) \\
e^{-i\theta X} &= \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix} = M(e^{-i\theta X}, \{|+\rangle, |-\rangle\}) \\
e^{-i\theta Y} &= \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix} = M(e^{-i\theta Y}, \{|i\rangle, |-i\rangle\}).
\end{aligned}$$

Cada una en la base que diagonaliza la aplicación. Pasemos $e^{i\theta X}$ y $e^{i\theta Y}$ a la base usual:

$$\begin{aligned}
e^{-i\theta X} &= \frac{1}{2}\begin{pmatrix} e^{-i\theta} + e^{i\theta} & e^{-i\theta} - e^{i\theta} \\ e^{-i\theta} - e^{i\theta} & e^{-i\theta} + e^{i\theta} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -i\sin(\theta) \\ -i\sin(\theta) & \cos(\theta) \end{pmatrix} = M(e^{-i\theta X}, \{|0\rangle, |1\rangle\}) \\
e^{-i\theta Y} &= \frac{1}{2}\begin{pmatrix} e^{-i\theta} + e^{i\theta} & i(-e^{-i\theta} + e^{i\theta}) \\ i(e^{-i\theta} - e^{i\theta}) & e^{-i\theta} + e^{i\theta} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = M(e^{-i\theta Y}, \{|0\rangle, |1\rangle\}).
\end{aligned}$$

A partir de las aplicaciones anteriores definimos.

Definición 2.22. Definimos los operadores rotacionales como:

$$R_X(\gamma) = e^{-i\gamma\frac{X}{2}} = \begin{pmatrix} \cos(\frac{\gamma}{2}) & -i\sin(\frac{\gamma}{2}) \\ -i\sin(\frac{\gamma}{2}) & \cos(\frac{\gamma}{2}) \end{pmatrix}, \; R_Y(\gamma) = e^{-i\gamma\frac{Y}{2}} = \begin{pmatrix} \cos(\frac{\gamma}{2}) & -\sin(\frac{\gamma}{2}) \\ \sin(\frac{\gamma}{2}) & \cos(\frac{\gamma}{2}) \end{pmatrix}$$

$$R_Z(\gamma) = e^{-i\gamma\frac{Z}{2}} = \begin{pmatrix} e^{-i\frac{\gamma}{2}} & 0 \\ 0 & e^{i\frac{\gamma}{2}} \end{pmatrix}.$$

Observación 2.20. Las aplicaciones anteriores son isometrías lineales y hermitianas.

Veamos porque se denominan operadores rotacionales.

Proposición 2.13. *Si tomamos Φ la proyección de los qubits a la esfera de Bloch, $\Phi \circ R_X(\gamma) \circ \Phi^{-1}$ (respectivamente $R_Y(\gamma)$, $R_Z(\gamma)$) es un giro de ángulo γ con eje de rotación el eje X (respectivamente eje Y, eje Z).*

Demostración. Primero, $\Phi \circ R_X(\gamma) \circ \Phi^{-1}$ define (induce) una isometría lineal en $\mathbb{R}^3$, lo que vamos a hacer es comprobar que dicha isometría es un giro mirando los puntos fijos.

Sabemos que $|+\rangle = R_X(\gamma)|+\rangle$ y $|-\rangle = -R_X(\gamma)|-\rangle$, por tanto $\Phi|+\rangle$ y $\Phi|-\rangle$ son puntos fijos de la aplicación $\Phi \circ R_X(\gamma)$. Como es una isometría lineal con dos puntos fijos, entonces es la identidad o una simetría cuyo plano de simetría contenga al eje X o un giro respecto al eje X. Veamos que no es ni la identidad ni una simetría. Para ello tomaremos la imagen de

$|0\rangle$ y $|i\rangle$

$$\begin{aligned}
\Phi|0\rangle &= (0,0,1)\ ,\ \Phi|i\rangle = (0,1,0)\\
R_X(\gamma)|0\rangle &= \cos(\frac{\gamma}{2})|0\rangle + \sin(\frac{\gamma}{2})e^{-i\frac{\pi}{2}}|1\rangle\\
R_X(\gamma)|i\rangle &= \cos(\frac{\gamma-\frac{\pi}{2}}{2})|0\rangle + \sin(\frac{\gamma-\frac{\pi}{2}}{2})e^{-i\frac{\pi}{2}}|1\rangle\\
\Phi R_X(\gamma)|0\rangle &= (0,-\sin(\gamma),\cos(\gamma))\\
\Phi R_X(\gamma)|i\rangle &= (0,-\sin(\gamma-\frac{\pi}{2}),\cos(\gamma-\frac{\pi}{2})).
\end{aligned}$$

Si $\gamma \neq 2k\pi$, $k \in \mathbb{Z}$, entonces $\Phi|0\rangle \neq \Phi R_X(\gamma)|0\rangle$, $\Phi|i\rangle \neq \Phi R_X(\gamma)|i\rangle$ así que descartamos la identidad. Descartemos que sea una simetría:

$$\begin{aligned}
\Phi|0\rangle - \Phi R_X(\gamma)|0\rangle &= (0,\sin(\gamma),1-\cos(\gamma))\\
\Phi|i\rangle - \Phi R_X(\gamma)|i\rangle &= (0,1+\sin(\gamma-\frac{\pi}{2}),-\cos(\gamma-\frac{\pi}{2})) = (0,1-\cos(\gamma),-\sin(\gamma))
\end{aligned}$$

Si fuese una simetría $(0,\sin(\gamma),1-\cos(\gamma))$ y $(0,1-\cos(\gamma),-\sin(\gamma))$ tendrían que ser proporcionales pero no lo son para ningún γ. Por descarte $\Phi \circ R_X(\gamma) \circ \Phi^{-1}$ es un giro respecto al eje X. Para obtener el ángulo de giro basta con ver el ángulo que forma $\Phi|0\rangle$ con $\Phi R_X(\gamma)|0\rangle$, que es

$$\arccos(<\Phi|0\rangle, \Phi R_X(\gamma)|0\rangle>) = \arccos(\cos(\gamma)) = \gamma.$$

Análogamente se prueba con $R_Y(\gamma)$ y $R_Z(\gamma)$.

□

Por tanto los operadores rotacionales reciben su nombre porque rotan los qubits en la esfera de Bloch. A continación probemos que podemos ver toda puerta lógica como un giro en la esfera de Bloch.

Teorema 2.4 (Descomposición Z-Y). *Dada una puerta lógica U, existen $\alpha,\beta,\gamma,\delta \in \mathbb{R}$ tales que:*

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

Demostración. Expresemos U matricialmente como matriz unitaria

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} r_a e^{i\theta_a} & r_b e^{i\theta_b} \\ r_c e^{i\theta_c} & r_d e^{i\theta_d} \end{pmatrix}.$$

Como U es unitaria, entonces $U^\dagger = U^{-1}$

$$U^\dagger = \begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix} = \frac{1}{\det(U)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = U^{-1},$$

igualando componente a componente

$$\det(U)\overline{a} = d \ , \ \det(U)\overline{c} = -b \ , \ \det(U)\overline{b} = -c.$$

Y como $\det(U) = e^{i\theta}$ entonces

$$U = \begin{pmatrix} a & -b \\ \overline{b}e^{i\theta} & \overline{a}e^{i\theta} \end{pmatrix} = \begin{pmatrix} r_a e^{i\theta_a} & -r_b e^{i\theta_b} \\ r_b e^{i(\theta-\theta_b)} & r_a e^{i(\theta-\theta_a)} \end{pmatrix}.$$

Además $r_a^2 + r_b^2 = 1$, por tanto existe $\gamma \in \mathbb{R}$ tal que

$$U = e^{i\frac{\theta}{2}} \begin{pmatrix} \cos(\frac{\gamma}{2})e^{i(\theta_a - \frac{\theta}{2})} & -\sin(\frac{\gamma}{2})e^{i(\theta_b - \frac{\theta}{2})} \\ \sin(\frac{\gamma}{2})e^{i(\frac{\theta}{2} - \theta_b)} & \cos(\frac{\gamma}{2})e^{i(\frac{\theta}{2} - \theta_a)} \end{pmatrix}.$$

Denominando $\alpha = \frac{\theta}{2}$, $\beta = -(\theta_a + \theta_b)$ y $\delta = \theta_b - \theta_a$, obtenemos que

$$U = e^{i\alpha} R_Z(\beta) R_Y(\gamma) R_Z(\delta)$$

□

Observación 2.21. Toda puerta lógica es composición de R_Z y R_Y, por tanto, en la esfera de Bloch es un giro, ya que la composición de giros es un giro si los ejes se cortan..

Vamos a listar algunas de las puertas lógicas más importantes para un qubit:

- Las puertas de Pauli 2.21.
- Las puertas de rotación 2.22, de las cuales destacamos:
  - $S = R_Z(\frac{\pi}{2})$.
  - $T = R_Z(\frac{\pi}{4})$.
- Puerta de cambio de fase discreta, R_k:

$$R_k = \begin{pmatrix} 1 & 0 \\ 0 & e^{\pi i 2^{1-k}} \end{pmatrix}. \tag{2.3}$$

- $\sqrt{\text{NOT}}$:

$$\frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}.$$

- La puerta de Hadamard,H o $H^{\otimes 1}$, una de las más importantes:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

2.4.2. Varios qubits

Como indicamos al inicio de esta sección, las puertas lógicas son isometrías lineales. Si tenemos en cuenta que cuando operamos con n qubits, trabajamos en el espacio $\mathbb{C}^{2^n}$. Entonces una puerta lógica en la que participen n qubits es una isometría lineal de $\mathbb{C}^{2^n}$ en $\mathbb{C}^{2^n}$.

Vamos a destacar las puertas controladoras: aquellas que mirando uno de los qubits aplican otra puerta al resto. Veamos las más populares:

- Controlled Not (CNOT o CX), aplicado a la base usual $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$:

$$CX|00\rangle = |00\rangle \ , \ CX|01\rangle = |01\rangle \ , \ CX|10\rangle = |11\rangle \ , \ CX|11\rangle = |10\rangle .$$

  Si lo vemos matricialmente

$$CX = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

- Análogamente a CNOT, podemos definir CY , CZ y CH:

$$CY = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \ , \ CZ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \ , \ CH = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}.$$

- En tres qubits destacamos, CCNOT aplica X al último qubit si los dos primeros están en $|11\rangle$, y CSWAP intercambia el valor de los dos últimos bits si el primero está a $|1\rangle$. CSWAP no viola el principio de no clonación, intercambia el valor, no copia. Matricialmente:

$$CCNOT = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \ , \ CSWAP = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

En general si tenemos una puerta lógica para n qubits , U, podemos hacer la puerta controladora U, CU, como

$$CU = \left(\begin{array}{c|c} I & 0 \\ \hline 0 & U \end{array}\right).$$

Además de las puertas controladoras vamos a destacar las transformaciones de Walsh-Hadamard. Se definen recursivamente a partir de la puerta de Hadamard:

$$W_1 = H = H^{\otimes 1} \ , \ W_n = H \otimes W_{n-1} = H^{\otimes n}.$$

Estas puertas se aplican a n qubits y son relevantes para crear estados entrelazados. Si aplicamos al sistema $|00\rangle$ la puerta $H^{\otimes 2}$, obtenemos un estado entrelazado

$$H|00\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}.$$

3. Circuitos cuánticos

Daremos una introducción a los diagramas de circuitos cuánticos, una herramienta para visualizarlos. Además veremos como implementar dichos circuitos en el simulador qiskit.

3.1. Diagramas

Hemos descrito las componentes necesarias para un circuito cuántico que son:

- Su unidad de información, los qubits.
- Trabajar con la unidad de información, las puertas lógicas cuánticas u operadores.
- Leer el contenido de la unidad de información, las mediciones.

Con estos elementos ya estamos preparados para describir algoritmos y aplicaciones de computación cuántica las cuales veremos en el siguiente capítulo y en la segunda parte. Dar una descripción con solo qubits, operaciones y mediciones es díficil de seguir, por ello vamos a introducir los diagramas de circuitos cuánticos. Estos diagramas muestran gráficamente los qubits, las operaciones que realizamos con ellos, y las mediciones. Son útiles para de un vistazo comprender el funcionamiento del circuito o al menos el orden de ejecución de las puertas lógicas.

Como hemos indicado un circuito consta de tres elementos, veamoslos todos combinados en él diagrama 3.1.

Figura 3.1.: Ejemplo básico de un diagrama. A la izquierda un bit, en el medio una puerta lógica y a la derecha una medición.

El elemento a la izquierda es el qubit que entra al circuito en el estado $|\psi\rangle$, en la mayoría de los circuitos la entrada son qubits en el estado $|0\rangle$. El elemento intermedio representa a una puerta lógica, concretamente aplica el operador A. Finalmente, el elemento de la derecha representa una medición en la base computacional.

Ahora podemos complicar el diagrama añadiendo más elementos. En el diagrama 3.2 tenemos uno más complejo donde además hemos añadido puertas controladoras.

Cuando en la misma columna aparecen varias puertas se aplican dichas puertas a los qubits indicados mientras que al resto se le aplica la identidad formalmente. Por ejemplo la puerta A es un operador de C^4 en C^4 aplicada a los dos primeros quibts, mientras que B va de C^8 en C^8 y se aplica al cuarto, quinto y sexto qubits.

Por otro lado las puertas C y D son puertas controladoras, los qubits controladores se indican con un punto que luego conecta con la puerta. En este caso la puerta D está controlada por el tercer, sexto y séptimo qubits.

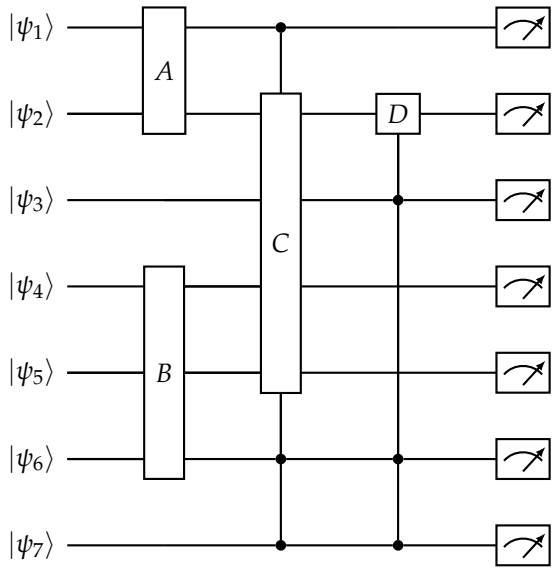


Figura 3.2.: Un diagrama más complejo. Aplicamos la puerta $A \otimes I \otimes B \otimes I$, luego una puerta C a los quibts segundo, tercero y cuarto controlada por el resto de qubits. Seguidamente se aplica el operador D al segundo qubit controlado por el tercer, sexto y séptimo qubits, al resto de qubits se le aplica la identidad. Finalmente medimos todos los qubits.

Si tenemos varios qubits a los cuales vamos a aplicar las mismas puertas, o para dejar diagramas más compactos, podemos agrupar los qubits en el diagrama. Ilustramos como se denota en 3.3.

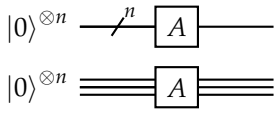


Figura 3.3.: Diagramas de dos circuitos, ambos hacen lo mismo pero cada uno representa una forma distinta de agrupar varios qubits a la vez.

Esbocemos los diagramas de las puertas lógicas que hemos visto en el capítulo anterior. En la figura 3.4 tenemos las puertas lógicas aplicadas a un solo qubit, mientras que, en la figura 3.5 tenemos ejemplos de puertas controladoras y de las puertas de Wals-Hadamard.

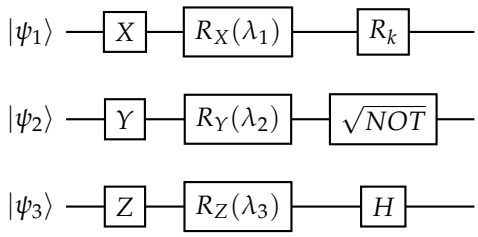


Figura 3.4.: Diagrama donde aparecen las puertas ya descritas aplicadas a un solo qubit.

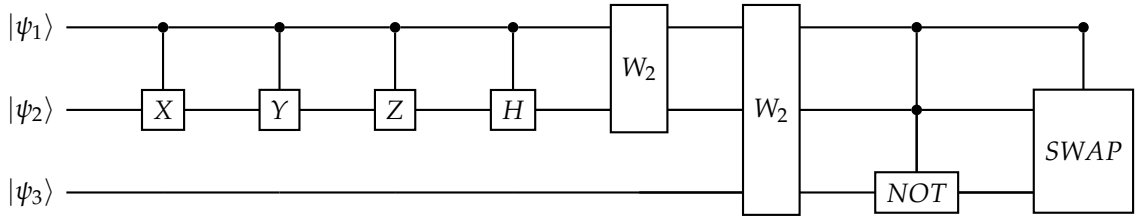


Figura 3.5.: Diagrama donde aparecen las puertas lógicas que hemos listado aplicadas a varios qubits. Primero aparecen las controladas por un solo qubit, a continuación algunas puertas de Wals-Hadamard y finalmente puertas controladoras aplicadas a tres qubits, respectivamente CCNOT y CSWAP.

También, a estos diagramas se le puede añadir bits clásicos e indicar en que bit estamos guardando la información de una medición. Otra utilidad que tienen los bits clásicos es interpretar el papel de bits de control en puertas lógicas, si un bits está a 0, no se ejecuta la puerta, y si está a 1, se opera. Cuando metemos este tipo de condición, la puerta lógica deja de ser reversible. Un ejemplo es el diagrama 3.6. La doble línea indica que estamos guardando la medición en ese bit, y, la forma de controlar es igual que con qubits con la diferencia de que si el 0 es el que permite la operación se indica.

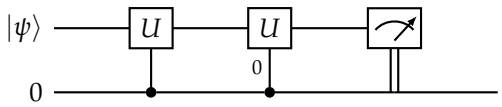


Figura 3.6.: Diagrama con qubits y bits. Dos controladoras, una que permite la acción del operador si el bit está a 1 y la segunda si está a 0. La doble línea indica que se esta guardando la medición en el bit.

3.2. Qiskit

Con los diagramas tenemos herramientas para describir circuitos cuánticos de forma gráfica. Pero aún seguimos en un terreno teórico, no hemos presentado ninguna forma de llevarlo a la práctica. Dado que actualmente los computadores cuánticos están en desarrollo, solo se han podido construir con pocos qubits, es más usual usar librerías que simulen el comportamiento de un ordenador cuántico en uno clásico. Hay diversas librerias para este fin,

como por ejemplo son Cirq de Google y Pennylane, ambas de software libre y disponibles en Python. No vamos a usar ninguna de las dos librerias anteriores, nostros hemos optado por el simulador Qiskit (Quantum Information Software Kit for Quantum Computation) de IBM que también es de software libre y está en el lenguaje Python. El motivo de elegir este simulador sobre otros es porque está mejor documentado y tiene una buena retrocompatibilidad con versiones más antiguas.

Vamos a introducir como se implementan los circuitos en Qiskit, esta información se ha obtenido de la documentación oficial disponible en [Qis23]. Podemos separar los elementos en 5 componentes, los registros, los operadores, el circuito, los medidores y el simulador. Examaminemos cada componente:

- Registros. Hay de dos tipos clásicos y cuánticos, cada uno contiene bits y qubits respectivamente. Para crear un registro clásico se puede usar la función ClassicalRegister(n) y para qubits QuantumRegister(n), donde n indica el número de bits o qubits que queremos en el registro.
- Circuitos. Se crean a partir de los registros con la función QuantumCircuit(rq, rc), donde rq y rc son registros cuánticos y clásicos respectivamente. No es necesario incluirlos en la declaración del circuito se puede añadir posteriormente, o incluso sustituirlos por naturales e implicitamente se crean los registros. A un circuito le añadimos los operadores y medidores. También podemos juntar varios circuitos en uno y hacer subcircuitos.
- Operadores. Añadir los operadores a un circuito, también podemos convertir un circuito en un operador si en este solo intervienen qubits y operadores. El esquema general para añadir operador de un solo qubit es circ.gate([,]n). Donde circ es el circuito, n indica a que qubit se le está aplicando la puerta, gate es un tipo de operador y [,] serían parametros correspondientes al operador. Algunos operadores de un solo bit que tiene qiskit son los operadores de pauli (x,y,z,id), de rotacion (rz(λ), ry(λ), rz(λ)), Hadamard (h) y operador arbitario U (u(β, γ, δ) donde estos parámetros son los correspondientes a la descomposición ZY). También tiene puertas controladoras por qubits, indicando primero los qubits que controlan y luego a cuales se aplican, y puertas controladoras por bits.
- Los medidores miden un qubit y dan la medición en un bit.
- Simulador. Como indica su nombre se encargan de dado la entrada del circuito calcular los resultados. Se puede simular multiples veces el circuito pudiendo obtener datos estadísticos.

4. Algunas aplicaciones

En este capítulo veremos dos aplicaciones muy sencillas de la computación cuántica, superdense coding y teleportation. Después pasaremos a describir el algoritmo de Shor para el que necesitaremos la transformada de Fourier cuántica y el algoritmo de estimación de fase.

4.1. Superdense coding

La primera aplicación que vamos a ver es que un qubit puede transmitir la información de dos bits, para ello vamos a necesitar un par ERP, $|\psi_0\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$, y supongamos que tenemos dos personas Alice y Bob. Alice tiene un qubit del ERP y Bob el otro.

Supongamos que Alice tiene dos bits, dependiendo de la codificación de estos aplicará una puerta de Pauli a su qubit, como el qubit de Bob no interviene, a este se le aplica la identidad. Veamoslo en la tabla 4.1.

Bits	Puerta	Resultado
00	$\|\psi_0\rangle = \sigma_0 \otimes I \|\psi_0\rangle = I \otimes I \|\psi_0\rangle$	$\|\psi_0\rangle = \frac{\|00\rangle + \|11\rangle}{\sqrt{2}}$
01	$\|\psi_1\rangle = \sigma_1 \otimes I \|\psi_0\rangle = X \otimes I \|\psi_0\rangle$	$\|\psi_1\rangle = \frac{\|10\rangle + \|01\rangle}{\sqrt{2}}$
10	$\|\psi_2\rangle = \sigma_2 \otimes I \|\psi_0\rangle = Y \otimes I \|\psi_0\rangle$	$\|\psi_2\rangle = i\frac{\|10\rangle - \|01\rangle}{\sqrt{2}}$
11	$\|\psi_3\rangle = \sigma_3 \otimes I \|\psi_0\rangle = Z \otimes I \|\psi_0\rangle$	$\|\psi_3\rangle = \frac{\|00\rangle - \|11\rangle}{\sqrt{2}}$

Tabla 4.1.: Operaciones de Alice en superdense coding

Tras operar Alice envía su qubit a Bob, este último aplica CX al par ERP la cual deja de estar entrelazada. Tras ello realiza una medición respecto a la base usual o computacional en el segundo qubit forzando al primer qubit a tomar un valor. En la tabla 4.2 tenemos los resultados.

Estado inicial	CX	Primer qubit	Segundo qubit	Medición del 2ºqubit
$\|\psi_0\rangle = \frac{\|00\rangle + \|11\rangle}{\sqrt{2}}$	$\frac{\|00\rangle + \|10\rangle}{\sqrt{2}}$	$\frac{\|0\rangle + \|1\rangle}{\sqrt{2}}$	$\|0\rangle$	$\|0\rangle$
$\|\psi_1\rangle = \frac{\|10\rangle + \|01\rangle}{\sqrt{2}}$	$\frac{\|11\rangle + \|01\rangle}{\sqrt{2}}$	$\frac{\|1\rangle + \|0\rangle}{\sqrt{2}}$	$\|1\rangle$	$\|1\rangle$
$\|\psi_2\rangle = i\frac{\|10\rangle - \|01\rangle}{\sqrt{2}}$	$i\frac{\|11\rangle - \|01\rangle}{\sqrt{2}}$	$\frac{\|1\rangle - \|0\rangle}{\sqrt{2}}$	$i\|1\rangle$	$\|1\rangle$
$\|\psi_3\rangle = \frac{\|00\rangle - \|11\rangle}{\sqrt{2}}$	$\frac{\|00\rangle - \|10\rangle}{\sqrt{2}}$	$\frac{\|0\rangle - \|1\rangle}{\sqrt{2}}$	$\|0\rangle$	$\|0\rangle$

Tabla 4.2.: Operar CNOT y el valor del primer y segundo qubit tras medir

Finalmente aplicamos la puerta de Hadamard al primer qubit y lo medimos, en combinación con el segundo qubit nos queda una combinación única. En la tabla 4.3 están detalladas las operaciones.

Estado	Primer qubit	H primer qubit
$\|\psi_0\rangle$	$\frac{\|0\rangle + \|1\rangle}{\sqrt{2}}$	$\frac{1}{2}(\|0\rangle + \|1\rangle + \|0\rangle - \|1\rangle) = \|0\rangle$
$\|\psi_1\rangle$	$\frac{\|1\rangle + \|0\rangle}{\sqrt{2}}$	$\frac{1}{2}(\|0\rangle - \|1\rangle + \|0\rangle + \|1\rangle) = \|0\rangle$
$\|\psi_2\rangle$	$\frac{\|1\rangle - \|0\rangle}{\sqrt{2}}$	$\frac{1}{2}(\|0\rangle - \|1\rangle - \|0\rangle - \|1\rangle) = -\|1\rangle$
$\|\psi_3\rangle$	$\frac{\|0\rangle - \|1\rangle}{\sqrt{2}}$	$\frac{1}{2}(\|0\rangle + \|1\rangle - \|0\rangle + \|1\rangle) = \|1\rangle$

Tabla 4.3.: Aplicar H al primer qubit

Por tanto tras las correspondientes mediciones Bob conoce los dos bits que tenía Alice. Es más tras medir los qubits estos dan los bits de Alice. Para transmitir la información usamos un qubit pero para poder interpretarla usamos dos qubits.

4.2. Teleportation

Esta técnica permite reconstruir un qubit a partir de dos bits y un par ERP. Otra vez vamos a tener a Alice y Bob cada uno con un qubit del par ERP. Inicialmente Alice tendra un qubit $|\psi\rangle = a|0\rangle + b|1\rangle$. En combinación con el par ERP vamos a tener el siguiente estado:

$$\begin{aligned}|\psi\rangle \otimes |\psi_0\rangle &= (a|0\rangle + b|1\rangle) \otimes (\frac{|00\rangle + |11\rangle}{\sqrt{2}}) \\ &= \frac{a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle}{\sqrt{2}}.\end{aligned}$$

Empezamos aplicando $CX \otimes I$ y $H \otimes I \otimes I$ al estado anterior, obteniendo:

$$\begin{aligned}(H \otimes I \otimes I)(CX \otimes I)(|\psi\rangle \otimes |\psi_0\rangle) &= (H \otimes I \otimes I)(CX \otimes I)\frac{a|000\rangle + a|011\rangle + b|100\rangle + b|111\rangle}{\sqrt{2}} \\ &= (H \otimes I \otimes I)\frac{a|000\rangle + a|011\rangle + b|110\rangle + b|101\rangle}{\sqrt{2}} \\ &= \frac{a|000\rangle + a|100\rangle + a|011\rangle + a|111\rangle}{2} \\ &+ \frac{b|010\rangle - b|110\rangle + b|001\rangle - b|101\rangle}{2} \\ &= \frac{1}{2}(|00\rangle(a|0\rangle + b|1\rangle) + |10\rangle(a|0\rangle - b|1\rangle) \\ &+ |01\rangle(a|1\rangle + b|0\rangle) + |11\rangle(a|1\rangle - b|0\rangle))).\end{aligned}$$

Seguidamente Alice mide los dos primeros quibts, respecto a las proyecciones de los planos

generados por los vectores $\{|ij0\rangle, |ij1\rangle\}$ con $i, j \in \{0, 1\}$. La probabilidad de medir $|00\rangle$, $|01\rangle$, $|10\rangle$ y $|11\rangle$ es la misma. En la tabla 4.4 podemos ver como queda el qubit de Bob tras la medición de Alice ya que el estado colapsa.

Medición	Qubit de Bob
$\|00\rangle$	$a\|0\rangle + b\|1\rangle$
$\|01\rangle$	$a\|1\rangle + b\|0\rangle$
$\|10\rangle$	$a\|0\rangle - b\|1\rangle$
$\|11\rangle$	$a\|1\rangle - b\|0\rangle$

Tabla 4.4.: El qubit de Bob tras la medición de Alice

Alice envía su medición, que son bits, a Bob y por tanto él puede saber como está su par. Para recuperar el qubit $|\psi\rangle$ original Bob solo tiene que aplicar la correspondiente puerta lógica que será una de las siguientes $|\sigma_0, \sigma_1, \sigma_3, -i\sigma_2\rangle$ donde $-i\sigma_2$ es multiplicar por -i tras aplicar la puerta Y. Resumiendo en la tabla 4.5.

Bits	Puerta	Resultado
00	I	$I(a\|0\rangle + b\|1\rangle) = a\|0\rangle + b\|1\rangle = \|\psi\rangle$
01	X	$X(a\|1\rangle + b\|0\rangle) = a\|0\rangle + b\|1\rangle = \|\psi\rangle$
10	Z	$Z(a\|0\rangle - b\|1\rangle) = a\|0\rangle + b\|1\rangle = \|\psi\rangle$
11	-iY= ZX	$-iY(a\|1\rangle - b\|0\rangle) = -iia\|0\rangle + -iib\|1\rangle = \|\psi\rangle$

Tabla 4.5.: El qubit de Bob tras la medición de Alice

Finalmente reconstruimos el qubit original. Hay que recalcar que no se ha violado el principio de no clonacion 2.3, para clonar el qubit hemos tenido que medirlo por tanto lo hemos destruido además de que en el proceso es necesario usar un par ERP y medir uno de sus qubits. De esto podemos deducir que podemos guardar la información de un qubit en dos bits y luego recuperarla usando un par ERP.

Un diagrama que represente está aplicación sería el de la figura 4.1.

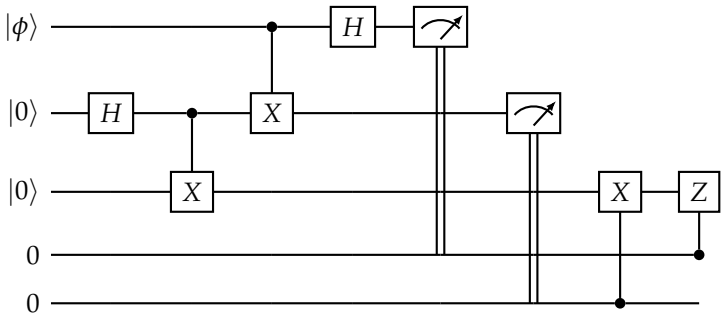


Figura 4.1.: Circuito de teleportation.

4.3. Algoritmo de Shor

En lo siguiente explicaremos el algoritmo de Shor que desmostró la utilidad de la computación cuántica en aplicaciones reales. Este algoritmo es capaz de descomponer un número N en tiempo polinomico, por tanto los sistemas de encriptación basados en descomposición de primos serían vulnerables en caso en construirse una computadora cuántica lo sufientemente potente. La única contraparte es que es probabilístico, pero ejecutándolo varias veces es plausible obtener el resultado deseado.

Para poder explicar el algoritmo necesitaremos de la transformada de Fourier y del algoritmo de estimación de fase cuánticos.

4.3.1. Transformada de Fourier cuántica

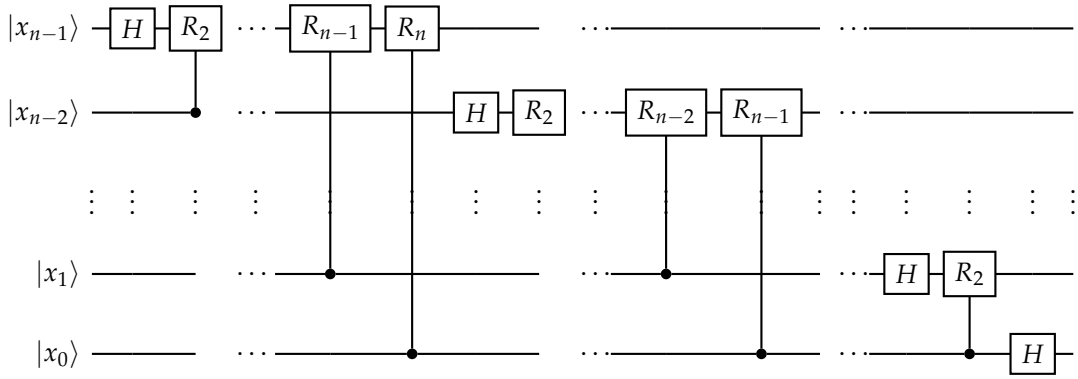
Sea un estado cuántico, $|\psi\rangle = |x_0 x_1 \dots x_{n-1}\rangle$ con $x_i \in \{0,1\}$, denotaremos

$$x = 2^{n-1}x_0 + 2^{n-2}x_1 + \cdots + 2^1 x_{n-2} + 2^0 x_{n-1} = \sum_{k=0}^{n-1} 2^{n-1-k} x_k.$$

Definición 4.1. Dado un estado $|\psi\rangle = |x_0 \dots x_{n-1}\rangle$, la transformada de Fourier se define como:

$$\begin{aligned}&\frac{1}{2^{n/2}}(|0\rangle + e^{\pi i \frac{x}{2^{n-1}}}|1\rangle) \otimes (|0\rangle + e^{\pi i \frac{x}{2^{n-2}}}|1\rangle) \otimes \cdots \otimes (|0\rangle + e^{\pi i \frac{x}{2^0}}|1\rangle)\\ &= \frac{1}{\sqrt{2}} \bigotimes_{j=0}^{n-1} (|0\rangle + e^{\pi i \frac{x}{2^{n-1-j}}}|1\rangle).\end{aligned}$$

La transformada de Fourier se puede realizar operando con puertas lógicas simples, concretamente puertas de Hadamard, H, y puertas de cambio de fase discreta controladoras (2.3), CR_k. Dado $|\psi\rangle = |x_{n-1} \dots x_0\rangle$, empezando en $i = n-1$ y acabando en $i = 0$, primero aplicamos H a x_i, después desde $j = i-1$ hasta $j = 0$ operamos CR_{1+i-j} a x_i controlado por x_j. En un diagrama:



Si denominamos como F a la aplicación que asigna a cada vector su transformada de Fourier, por la descomposición anterior F, es una aplicación lineal unitaria al ser producto tensorial de aplicacicones unitarias y composición de unitarias. Así que podemos hablar de la inversa de la transformada de Fourier, que como aplicación sería $F^\dagger$.

4.3.2. Algoritmo estimación de fase cuántica

Dada una puerta lógica U podemos diagonalizarla, porque es unitaria, por el teorema de diagonalización espectral. Además, como es unitaria, todos sus valores propios deben de ser complejos de modulo 1, por tanto, dado una valor propio λ_j de U, existe un $0 \leq \theta_j \leq 1$ tal que $\lambda_j = e^{2\pi i\theta_j}$. El algortimo de estimación de fase trata de dar los primeros m bits de θ_j.

Para el algoritmo necesitamos una puerta unitaria U y un vector propio $|\phi\rangle$ de U, buscamos estimar con m bits de precisión el ángulo θ asociado al valor propio del vector propio $|\phi\rangle$. Como entrada al algoritmo se tiene

$$|\phi_{in}\rangle = |0\rangle^{\otimes m} |\phi\rangle .$$

A todos los $|0\rangle$ les aplicamos H, obteniendo

$$|\phi_1\rangle = (H^{\otimes m} \otimes I) |\phi_{in}\rangle = \frac{1}{\sqrt{2^n}}(|0\rangle + |1\rangle)^{\otimes m} |\phi\rangle .$$

Se aplican puertas controladoras CU^{2^j} con $0 \leq j \leq m-1$ a $|\phi\rangle$ controladas cada una por un $H|0\rangle$ distinto. Antes de expresar el resultado, recalcamos que

$$\begin{aligned} U^{2^j} |\phi\rangle &= U^{2^{j-1}} e^{2\pi i\theta} |\phi\rangle = \cdots = e^{2\pi i\theta 2^j} |\phi\rangle , \\ CU^{2^j}(|0\rangle + |1\rangle) \otimes |\phi\rangle &= |0\rangle \otimes |\phi\rangle + |1\rangle \otimes e^{2^{j+1}\pi i\theta} |\phi\rangle = (|0\rangle + e^{2^{j+1}\pi i\theta} |1\rangle) \otimes |\phi\rangle \\ &= (|0\rangle + e^{2^{j+1}\pi i\theta} |1\rangle) |\phi\rangle . \end{aligned}$$

Tras aplicar los operadores se obtiene

$$\begin{aligned} |\phi_2\rangle &= \frac{1}{\sqrt{2^m}} \left(|0\rangle + e^{\mathbf{2\pi i}\theta 2^{m-1}}|1\rangle\right) \otimes \cdots \otimes \left(|0\rangle + e^{\mathbf{2\pi i}\theta 2^1}|1\rangle\right) \otimes \left(|0\rangle + e^{\mathbf{2\pi i}\theta 2^0}|1\rangle\right) \otimes |\phi\rangle \\ &= \frac{1}{\sqrt{2^m}} \sum_{k=0}^{2^m-1} e^{\mathbf{2\pi i}\theta k}|k\rangle \otimes |\phi\rangle. \end{aligned}$$

En la anterior sumatoria aparece un $|k\rangle$ con $0 \leq k \leq 2^{m-1}$, esto denota el $|a\rangle$ donde a es la codificación de k en binario utilizando m bits. Por ejemplo, para $m = 3$, $k = 0$, $k = 5$ y $k = 7$ con $m = 3$ serían:

$$|0\rangle = |000\rangle \ , \ |5\rangle = |101\rangle \ , \ |7\rangle = |111\rangle .$$

Ahora vamos a prescindir de $|\phi\rangle$ y trabajaremos con la otra parte del producto tensorial. Si consideramos a $|2^m\theta\rangle$ vemos que se ha calculado su fransformada de Fourier en $|\phi_2\rangle$, por tanto si aplicamos la inversa recuperariamos dicho estado

$$|\phi_3\rangle = \frac{1}{2^m} \sum_{x=0}^{2^m-1} \sum_{k=0}^{2^m-1} e^{-\frac{2\pi ik}{2^m}(x-2^m\theta)} |x\rangle \otimes |\phi\rangle .$$

Aproximamos $2^m\theta$ por el entero más cercano, entonces existen a y δ tales que $2^m\theta = a + 2^m\delta$

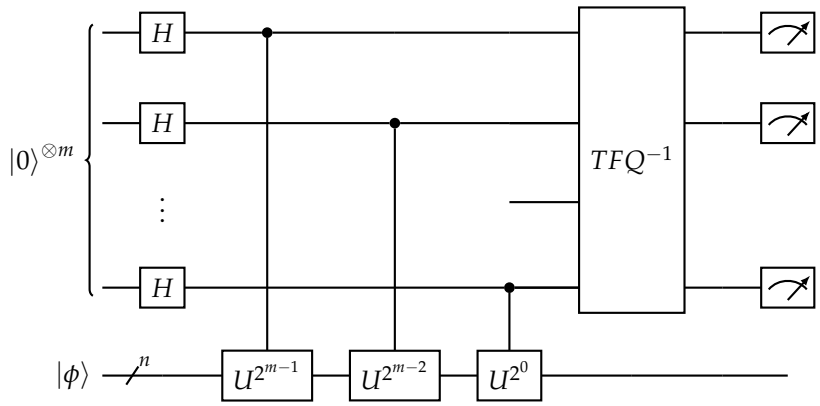


Figura 4.2.: Circuito cuántico del algoritmo de estimación de fase cuántica

y con $0 \leq |2^m\delta| \leq 1$. Agregando esto

$$|\phi_3\rangle = \sum_{x=0}^{2^m-1} \frac{1}{2^m} \sum_{k=0}^{2^m-1} e^{2^{1-m}\pi ik(x-a)} e^{2^m\delta} |x\rangle \otimes |\phi\rangle .$$

El último paso es medir los primeros m qubits de $|\phi_3\rangle$. En la figura 4.2 tenemos un circuito que representa el funcionamiento del algoritmo.

Como las mediciones tienen asociada una probabilidad, este algoritmo tiene un carácter. Veamos la probabilidad de medir $|a\rangle$ o a, que es la aproximación que estamos buscando

$$Pr(a) = |\frac{1}{2^m} \sum_{k=0}^{2^m-1} e^{2^k\delta}|^2 = \frac{1}{2^{2m}} |\sum_{k=0}^{2^m-1} e^{2^k\delta}|^2.$$

Si $\delta = 0$, la probabilidad es 1 y por tanto el algoritmo es exacto. En otro caso sabiendo que la serie anterior es una geometrica con razón $e^{2\pi i\delta}$

$$\begin{aligned}
Pr(a) &= \frac{1}{2^{2m}} |\frac{1 - e^{2^m\pi i\delta}}{1 - e^{2\pi i\delta}}|^2 \ , \ \delta \neq 0 \\
&= \frac{1}{2^{2m}} |\frac{\sin(\pi 2^m\delta)}{\sin(\pi\delta)}|^2 \ , \ |1 - e^{2ix}|^2 = |1 - \cos^2(x) - 2i\cos(x)\sin(x) + \sin^2(x)|^2 = 4|\sin(x)|^2 \\
&\geq \frac{1}{2^{2m}} |\frac{\sin(\pi 2^m\delta)}{\pi\delta}|^2 \\
&\geq \frac{1}{2^{2m}} |\frac{22^m}{\pi\delta}|^2 \\
&\geq \frac{4}{\pi^2} \approx 0.40528 \ldots .
\end{aligned}$$

Esta probabilidad se suele considerar suficiente para varios contextos, por ejemplo en el algoritmo de Shor. Igualmente si se desea, se puede aumentar la probabilidad hasta $1 - \epsilon$, añadimos $O(\log(\frac{1}{\epsilon}))$ qubits y tras la medición los ignoramos, está detallado en [CEMM98].

Por otro lado, qué ocurriría si tomamos un qubit $|\psi\rangle$ arbitrario. Por el teorema de descomposición espectral sabemos que podemos expresar $|\psi\rangle$ como combinación lineal de vectores propios y ortonormales:

$$|\psi\rangle = \sum_{h=1}^{n} \lambda_h e^{2\pi i \theta_h} |\psi_h\rangle .$$

Si aplicamos el algoritmo hasta antes de la medición, por linealidad, estariamos haciendo los cálculos para todas los vectores propios a la vez, quedaría la siguiente expresión:

$$\sum_{h=1}^{n} \sum_{x=0}^{2^m-1} \frac{1}{2^m} \sum_{k=0}^{2^m-1} \lambda_h e^{-\frac{2\pi i k}{2^m}(x-2^m\theta_h)} |x\rangle \otimes |\psi_h\rangle .$$

Cuando midamos podremos obtener alguna de todas las fases que se verá influenciada por $|\lambda_h|^2$ la probabilidad de medir ese frente a otra. Esto aparentemente no es útil pero la clave está en buscar algo general que tengan todos los posibles resultados anteriores.

4.3.3. Algoritmo de Shor

El algoritmo de Shor calcula un factor primo de un número impar N dado. En caso de que no encuentre, N es primo. Nosotros vamos a detallar cómo se puede realizar a través del algortimo de estimación de fase.

La idea consiste en dado $1 < x < N$, encontrar su periodo, es decir, el mínimo natural r tal que $x^r = 1 \ (mod \ N)$. Además, si r es par entonces $(x^{\frac{r}{2}} + 1)(x^{\frac{r}{2}} - 1) = 0 \ (mod \ N)$ y se tendría que m.c.d$((x^{\frac{r}{2}} + 1), N)$ o m.c.d$((x^{\frac{r}{2}} - 1), N)$ es un factor propio de N, donde m.c.d denota al máximo común divisor.

Describamos primero como hallar dicho periodo r dado un x que es la parte cuántica del algoritmo. Definimos el siguiente operador unitario

$$U_x |y\rangle = |xy \bmod N\rangle ,$$

los vectores propios de está puerta son, con $0 \leq t \leq r-1$,

$$|u_t\rangle = \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} e^{-2\pi i \frac{tk}{r}} |x^k \bmod N\rangle$$

con valores propios

$$\begin{aligned}
U_x |u_t\rangle &= \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} e^{-2\pi i \frac{tk}{r}} U_x |x^k \bmod N\rangle \\
&= \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} e^{-2\pi i \frac{tk}{r}} |x^{k+1} y \bmod N\rangle \\
&= \frac{1}{\sqrt{r}} (e^{-2\pi i \frac{t(r-1)}{r}} |x^r y \bmod N\rangle + \sum_{k=1}^{r-1} e^{-2\pi i \frac{t(k-1)}{r}} |x^k y \bmod N\rangle) \\
&= \sqrt{r} (e^{-2\pi i \frac{t(r-1)}{r}} |y \bmod N\rangle + \sum_{k=1}^{r-1} e^{-2\pi i \frac{t(k-1)}{r}} |x^k y \bmod N\rangle) \\
&= \frac{1}{\sqrt{r}} e^{2\pi i \frac{t}{r}} (e^{-2\pi i \frac{tr}{r}} |x^r y \bmod N\rangle + \sum_{k=1}^{r-1} e^{-2\pi i \frac{tk}{r}} |x^k y \bmod N\rangle) \\
&= e^{2\pi i \frac{t}{r}} \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} e^{-2\pi i \frac{tk}{r}} |x^k \bmod N\rangle \\
&= e^{2\pi i \frac{t}{r}} |u_t\rangle .
\end{aligned}$$

Además verifican

$$\begin{aligned}
\frac{1}{\sqrt{r}} \sum_{t=0}^{r-1} |u_t\rangle &= \frac{1}{\sqrt{r}} \sum_{t=0}^{r-1} \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} e^{-2\pi i \frac{tk}{r}} |x^k \bmod N\rangle \\
&= \frac{1}{r} \sum_{k=0}^{r-1} (\sum_{t=0}^{r-1} e^{-2\pi i \frac{tk}{r}}) |x^k \bmod N\rangle \\
&= \frac{1}{r} \sum_{k=0}^{r-1} r |x^k \bmod N\rangle \\
&= |1\rangle .
\end{aligned}$$

Al $|1\rangle$ vamos a aplicarle el algortimo de estimación de fase cuántica. Como hemos comentado obtendremos una de las posibles fases, que en este caso son de la forma $\frac{t}{r}$.

Todas tienen en común a r como denominador. Para hallar r se usan otras técnicas como fracciones continuas. Por otro lado, llegados a este punto, se puede dar por concluida la computación cuántica y podemos continuar haciendo computación clásica. Es posible continuar con la primera pero dado que no se obtiene ningún beneficio, y en ciertos sentidos es más trabajoso, se prefiere usar la segunda.

Como se prometío, una vez explicada la parte cuántica del algoritmo de Shor, describamos un pseudocódigo 1 que realiza el algoritmo.

Algorithm 1 Algoritmo de Shor

Input: entero impar N

1: Sea $A = \{2, 3, \ldots, N-1\}$.
2: **while** $A \neq \emptyset$ **do**
3: Extraer pseudo-aleatoriamente $x \in A$.
4: **if** m.c.d.$(x, N)! = 1$ **then**
5: Fin, m.c.d.(x, N) es un factor propio de N.
6: **else**
7: Calcular r, el periodo de x. ▷ Parte Cuántica
8: Si r es impar, volver al paso 2.
9: Si $x^{\frac{r}{2}} = -1 \bmod N$, volver al paso 2.
10: Fin, m.c.d.$(x^{\frac{r}{2}} - 1, N)$ o m.c.d.$(x^{\frac{r}{2}} + 1, N)$ son factores propio de N.

Dos comentarios sobre el algoritmo 1. Respecto a la exigencia de r impar, la exigencia de r impar, el motivo es poder dividir por 2, y la exigencia de $x^{\frac{r}{2}} \neq -1 \bmod N$ es porque solo puede valer 1 o -1 y para aplicar el resultado es necesario que sea 1. La pregunta natural entonces es si r puede verificar ambas condiciones. De hecho sí y es bastante probable. Sea $N = p_1^{\alpha_1} \ldots p_k^{\alpha_k}$, tomando $1 < x \leq N-1$ aleatoriamente se verifica

$$Pr(r \text{ par } \wedge x^{\frac{r}{2}} \neq -1 \bmod N) \geq 1 - (\frac{1}{2})^k.$$

Finalmente, al inicio de la sección deciamos que el algoritmo de Shor era polinómico en tiempo. Se puede comprobar en su propio texto original [Sho97], pero se debe de implementar con cuidado ya que hay puertas lógicas cuyo cómputo puede ser muy deficiente. También en el texto original se comenta la veracidad de los resultados algebraicos usados.

Parte II.

Aprendizaje Automático Cuántico a Series Temporales

5. Aprendizaje Automático Cuántico

5.1. Definición

El aprendizaje automático (AA, en inglés Machine Learning, ML) es una dísciplina de la Inteligencia Artificial que busca a partir de la experiencia, dar modelos para resolver problemas. Más formalmente un modelo depende de la experiencia E con respecto a una clase de tareas T y medida de desempeño P, si

$$\uparrow\uparrow E \longrightarrow P_T \text{ también aumenta.}$$

Dentro del AA tenemos varios paradigmas:

- Aprendizaje supervisado: aprender una función objetivo con una muestra de datos y sus respuestas, por ejemplo máquinas de vectores de soporte.
- Aprendizaje no supervisado: aprender propiedades de los datos sin saber la respuesta de cada uno.
- Conocimiento de transferencia: cómo compartir conocimiento entre tareas.
- Aprendizaje por refuerzo: compuesto por datos y recompensas. Se obtiene una política de actuación que trata de maximizar las recompensas.

Combinando el aprendizaje automático clásico y la computación cuántica, tenemos el aprendizaje automático cuántico. Una primera clasificación que podemos hacer es la indicanda en la figura 5.1:

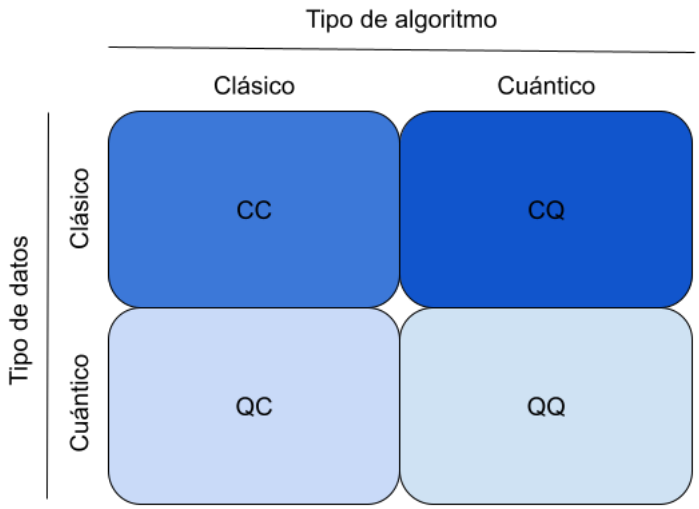


Figura 5.1.: Clasificación del aprendizaje automático cuántico según tipo de dato y algoritmo.

- CC, datos clásicos usando algoritmos clásicos. Es el aprendizaje automático clásico aplicado directamente en un computador cuántico.

- QC, datos cuánticos a los que aplicamos algoritmos clásicos. Tenemos un dato cuántico, lo medimos y aplicamos un algortimo tradicional.

- CQ, datos clásicos a los que aplicamos algoritmos cuánticos. El principal uso que tienen es para el minado de datos (data mining) usando algoritmos clásicos adaptados o algoritmos que usan propiamente las propiedades un ordenador cuántico. Distanciándonos del aprendizaje automático el algortimo de Shor puede caer en está categoría.

- QQ, datos cuánticos con algoritmos cuánticos. Usamos datos cuñanticos obtenidos tras un experimento o un algoritmo y los usamos como entrada de otro algoritmo. Un ejemplo alejándonos del aprendizaje automático sería aplicar el algoritmo de estimación de fase.

Actualmente los algoritmos cuánticos podemos clasificarlos en tres categorías:

- Algoritmos clasicos adaptados, por ejemplo máquinas de vectores de soporte cuánticas.

- Algoritmos puramente cuánticos, son de inspiración propiamente cuánticos o usan dichas propiedades potenciar algoritmos clásicos.

- Algoritmos híbiridos, combinan algoritmos tradicionales y cuánticos para obtener un mayor rendimiento y reducir el coste de aprendizaje

Nosotros nos vamos a centrar especialmente en CQ, ya que las series temporales son datos que tomamos y estos son clásicos. Hay dos formas de proceder, la primera es ejecutando algoritmos tradiconales en computadores cuánticos o simuladores con el objetivo de mejorar el rendimiento, y la segunda, es crear algoritmos cuánticos basadas en subrutinas cuánticas como el algoritmos de estimación de fase o las redes variacionales que veremos más tarde.

Igualmente para poder trabajar necesitamos transformar nuestros datos clásicos para poder aplicarles nuestros algoritmos cuánticos, a este proceso se le llama codificación cuántica (quatum encoding). La figura 5.2, pone en contraposición el cómo se realiza el AA tradicional y el AA cuántico en un ordenador clásico y otro cuántico. En un ordenador tradicional el proceso es el siguiente: dada una entrada, aplicamos el algoritmo y obtenemos resultados. Mientras que en uno cuántico es: dada un entrada, la codificamos, aplicamos el algoritmo cuántico, medimos y obtenemos resultados.

5.2. Codificación cuántica

Como se ha comentado antes, para trabajar con datos clásicos en un ordenador cuántico o simulador tenemos que codificarlos. La forma en la que codificamos da un contexto de como diseñar un algoritmo cuántico y las mejoras de rendimiento que podemos esperar, es más, la codificación de datos se puede considerar parte del algoritmo y puede ser una parte crucial de su complejidad.

Las codificaciones que vamos a tratar son: bases, amplitud, qsample y hamiltoniano.

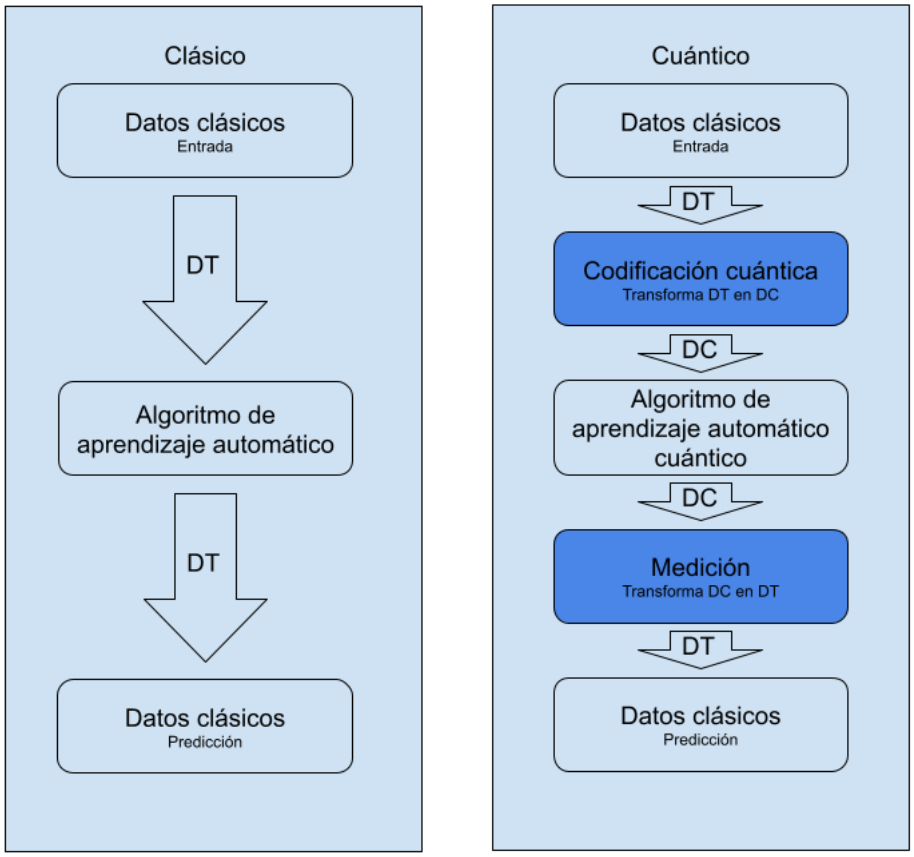


Figura 5.2.: A grandes rasgos funcionamiento del aprendizaje automático en ordenadores clasicos y cuánticos. DT denota datos clásicos o tradicionales mientras que DC son datos cuańticos.

5.2.1. Base

Dado un conjunto de datos binarios $\mathcal{D}$ de tamaño N donde cada $x^i \in \mathcal{D}$ es un vector binario de longitud M, $x^i = (b_1^i, \ldots, b_M^i)^T$ $b_j \in \{0,1\}$ $1 \leq j \leq M$ $1 \leq i \leq N$. La codificación en base de $\mathcal{D}$ es la superposición de los siguientes estados

$$|\mathcal{D}\rangle = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} |x^i\rangle .$$

Por ejemplo si $\mathcal{D} = \{(0,1,0,1)^T, (1,1,1,1)^T, (0,1,1,1)^T, (0,1,0,1)^T\}$, entonces la codificación base de $\mathcal{D}$ es

$$|\mathcal{D}\rangle = \frac{1}{\sqrt{4}} \sum_{i=1}^{4} |x^i\rangle = \frac{1}{\sqrt{4}}(|0101\rangle + |1111\rangle + |0111\rangle + |0101\rangle) = \frac{1}{\sqrt{4}}(2\,|0101\rangle + |1111\rangle + |0111\rangle).$$

Una implementación de esta codificación es usando M qubits que serán los registros de carga $\{l_1 \ldots, l_M\}$, 2 qubits auxiliares o ancillas $\{a_1, a_2\}$ y otros M qubits que serán los

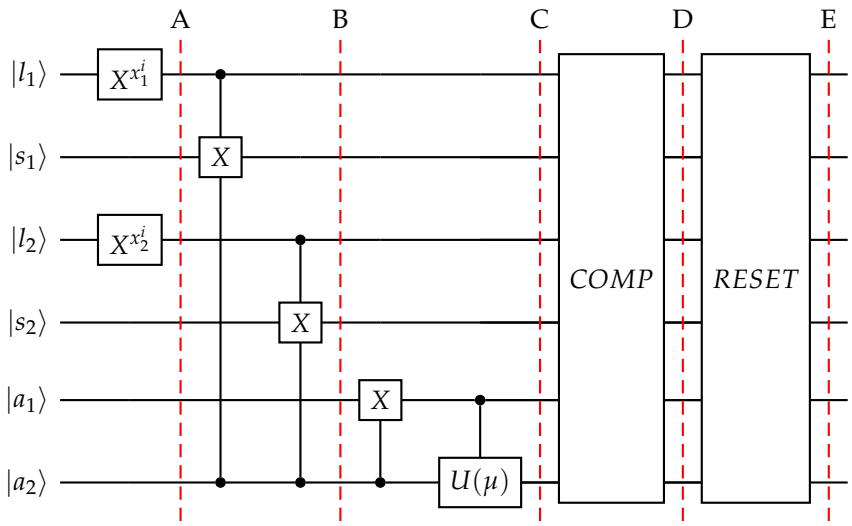


Figura 5.3.: Circuito de codificación base para conjuntos con $M = 2$. A carga la codificación de los elementos en los registros l, B copia l en s, C separa la parte de l y s que nos interesa, D suma, E restea los registros correspondientes.

registros de almacenamiento $\{s_1 \ldots, s_M\}$, en los registros s se guardará el resultado final. Lo notaremos de forma compacta como

$$|l_1 \ldots l_M : a_1 s_2 : s_1 \ldots s_M\rangle .$$

Lo primero dado el primer elemento $X^1 = (x_1^1, \ldots, x_M^1)$ definimos el siguiente estado

$$\frac{1}{\sqrt{N}} |0 \ldots 0 : 00 : x_1^1 \ldots x_M^1\rangle + \sqrt{\frac{N-1}{N}} |0 \ldots 0 : 01 : 0 \ldots 0\rangle .$$

El término con $a_2 = 0$ se denomina rama de memoria y la de $a_2 = 1$ rama de procesamiento. El algoritmo itera cargando en los registros de carga el valor de x^i, después lo "partez suma la parte correspondiente en los registros de almacenamiento, después restea los registros para la siguiente iteración.

Para explicar estas iteraciones, supongamos que llevamos i iteraciones realizadas y vamos a hacer la $i + 1$. Antes de ejecutar la iteración los registros están de la siguiente forma

$$|\phi^i\rangle = \frac{1}{\sqrt{N}} \sum_{k=1}^{i} |0 \ldots 0 : 00 : x_1^k \ldots x_M^k\rangle + \sqrt{\frac{N-i}{N}} |0 \ldots 0 : 01 : 0 \ldots 0\rangle .$$

Cargamos en los registros l el valor de $x^{(i+1)}$, a continuación usando un puerta NOT controlada por l_j y por a_2 pasamos el valor de l_j a s_j. Por tanto nos quedaría

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{i} |x_1^{(i+1)} \ldots x_M^{(i+1)} : 00 : x_1^k \ldots x_M^k\rangle + \sqrt{\frac{N-i}{N}} |x_1^{(i+1)} \ldots x_M^{(i+1)} \ldots 0 : 01 : x_1^{(i+1)} \ldots x_M^{(i+1)}\rangle .$$

Ahora mismo solo hemos cargado el valor de $x^{(i+1)}$, procedemos a prepararlo para sumar el

resultado. Aplicamos un NOT a a_1 controlado por a_2

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{i}|x_1^{(i+1)}\dots x_M^{(i+1)}:00:x_1^k\dots x_M^k\rangle+\sqrt{\frac{N-i}{N}}\,|x_1^{(i+1)}\dots x_M^{(i+1)}\dots 0:11:x_1^{(i+1)}\dots x_M^{(i+1)}\rangle\,.$$

Definimos $\mu = M+1-(i+1)$ y la puerta

$$U(\mu)=\begin{pmatrix}\sqrt{\frac{M-1}{\mu}} & \frac{1}{\sqrt{\mu}}\\ -\frac{1}{\sqrt{\mu}} & \sqrt{\frac{M-1}{\mu}}\end{pmatrix}.$$

Aplicamos $U(\mu)$ a a_2 controlado por a_1, obteniendo

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{i}|x_1^{(i+1)}\dots x_M^{(i+1)}:00:x_1^k\dots x_M^k\rangle+\frac{1}{\sqrt{N}}\,|x_1^{(i+1)}\dots x_M^{(i+1)}\dots 0:10:x_1^{(i+1)}\dots x_M^{(i+1)}\rangle$$
$$+\frac{\sqrt{N-(i+1)}}{\sqrt{N}}\,|x_1^{(i+1)}\dots x_M^{(i+1)}\dots 0:11:x_1^{(i+1)}\dots x_M^{(i+1)}\rangle\,.$$

Fijándonos en el factor que tiene $a_1a_2 = 10$, es justo lo que queremos sumar. Aplicando un NOT controlado por a_2 negado (se aplica NOT si $a_2 = 0$, y no si $a_2 = 1$)

$$\frac{1}{\sqrt{N}}\sum_{k=1}^{i+1}|x_1^{(i+1)}\dots x_M^{(i+1)}:00:x_1^k\dots x_M^k\rangle+\frac{\sqrt{N-(i+1)}}{\sqrt{N}}\,|x_1^{(i+1)}\dots x_M^{(i+1)}\dots 0:11:x_1^{(i+1)}\dots x_M^{(i+1)}\rangle\,.$$

Finalmente reseteamos los resgistros de l y ponemos $a_1 = 0$, la forma de hacerlo es aprovechando la reversibilidad de las puertas cuánticas. En este caso revertimos los CNOT usados hasta la puerta $U(\mu)$

5.2.2. Amplitud

Dado un conjunto de datos $\mathcal{D}$ de tamaño N, donde cada $x^i \in \mathcal{D}$ es un vector real de dimensión M. Definimos $\alpha_{\mathcal{D}}$ como

$$\alpha_{\mathcal{D}} = C_{\text{norm}}(x_1^1,\dots,x_M^1,\dots,x_1^N,\dots,x_M^N)^T.$$

Donde C_{norm} es una constante positiva tal que se verifica $\|\alpha_{\mathcal{D}}\|^2 = 1$. Es decir

$$C_{\text{norm}} = \|(x_1^1,\dots,x_M^1,\dots,x_1^N,\dots,x_M^N)^T\|^{-1}.$$

La codificación en amplitud de este conjunto es dado $p \geq log_2(NM)$

$$|\mathcal{D}\rangle = \sum_{i}^{2^p}\alpha_{\mathcal{D}_i}\,|i\rangle\,.$$

A cada entrada del vector $\alpha_{\mathcal{D}}$ le asociamos un elemento de la base computacional y lo multiplicamos por dicha entrada.

Otra forma de poder realizar esta codificación es ver $\alpha_{\mathcal{D}}$ como una matriz, es decir

$$A_{\mathcal{D}} = C_{\text{norm}} \begin{pmatrix} x_1^1 & x_2^1 & \dots & x_M^1 \\ x_1^2 & x_2^2 & \dots & x_M^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^N & x_2^N & \dots & x_M^N \end{pmatrix}.$$

Y la codificación sería

$$|\mathbb{D}\rangle = \sum_{i=1}^{N} \sum_{j=1}^{M} A_{\mathcal{D},ij} |i\rangle |j\rangle .$$

Si tomamos el mismo ejemplo usado en la codificación anterior $\mathcal{D} = \{(0,1,0,1)^T, (1,1,1,1)^T, (0,1,1,1)^T, (0,1,0,1)^T\}$ y la codificamos como primero hemos descrito

$$C_{\text{norm}} = \frac{1}{\|(0,1,0,1,1,1,1,1,0,1,1,1,0,1,0,1)^T\|} = \frac{1}{\sqrt{11}}$$
$$\alpha_{\mathcal{D}} = \frac{1}{\sqrt{11}}(0,1,0,1,1,1,1,1,0,1,1,1,0,1,0,1)^T.$$

Tenemos que $N = 4$, $M = 4$, $NM = 16$ y $\log_2 8 = 4$. Tomando $p = 4$ la codificación sería

$$\begin{aligned} |\mathcal{D}\rangle &= \sum_{i=1}^{2^4} \alpha_{\mathcal{D}_i} |i\rangle \\ &= \frac{1}{\sqrt{11}} |0001\rangle + \frac{1}{\sqrt{11}} |0011\rangle + \frac{1}{\sqrt{11}} |0100\rangle + \frac{1}{\sqrt{11}} |0101\rangle + \frac{1}{\sqrt{11}} |0110\rangle + \frac{1}{\sqrt{11}} |0111\rangle \\ &+ \frac{1}{\sqrt{11}} |1001\rangle + \frac{1}{\sqrt{11}} |1010\rangle + \frac{1}{\sqrt{11}} |1011\rangle + \frac{1}{\sqrt{11}} |1101\rangle + \frac{1}{\sqrt{11}} |1111\rangle . \end{aligned}$$

La principal ventaja de esta condificación es que usa un número de qubits logarítmico en función del conjunto de datos.

5.2.3. Qsample

Se puede considerar como un híbrido entre codificación base y codificación en amplitud. Dado un vector de amplitud $v = (v_1, \dots, v_{2^n})$, $\|v\|^2 = 1$ la codificación qsample sería

$$|\phi_v\rangle = \sum_{i=1}^{2^n} v_i |i\rangle .$$

Si vemos $\{v_1, \dots, v_{2^n}\}$ como un distribución discreta hemos codificado dicha distribución en un qubit.

Está codificación se basa en la analogía que hay entre distribuciones y qubits, por ejemplo veamos la relación entre el producto de dos distrubiciones y 2 qubits. Supongamos que tenemos 2 distribuciones, $v = (v_1, \dots, v_{2^n})$, $w = (w_1, \dots, w_{2^m})$. Y consideremos las codificaciones qsample de cada uno $|\phi_v\rangle$, $|\phi_w\rangle$, consideremos que trabajamos con ambos qubits a la vez, es

decir en su producto tensorial

$$|\phi_v\rangle \otimes |\phi_w\rangle = \sum_{i=1}^{2^n} \sum_{j=1}^{2^m} v_i w_j |i\rangle |j\rangle .$$

Si medimos este qubit, la probabilidad de $|i\rangle |j\rangle$ es $v_i w_j$ dada por la distribución producto de v y w.

En esencia está codificación traslada distribuciones a qubits, con los cuales podemos operar.

5.2.4. Hamiltoniano

Está codificación en vez ser directa como las anteriores es más ímplicita. Ahora dada una matriz hermítica, A, la codificamos para que defina el Halmiltoniano del sistema, H_A. El hamiltoniano es una aplicación lineal que regula la evolución del sistema en el tiempo de la siguiente manera

$$|\phi(t)\rangle = e^{-iH_A t} |\phi(0)\rangle .$$

El cálculo de la exponencial de una aplicación lineal es en general costoso, por tanto, un problema de la computación cuántica es el implementar la evolución dado un hamiltoniano. A dicho problema se le denomina simulación del hamiltoniano y formalmente se plantea: Dado un hamiltoniano H, un estado cuántico $|\phi\rangle$, un error $\epsilon > 0$, un t_0 y una norma $||||$, encontrar un algoritmo que implmente la evolución del sistema tal que el estado final $|\tilde{\phi}\rangle$ del algoritmo y $|\phi'\rangle = e^{-iH_A t_0} |\phi\rangle$ están a una distancia menor que ϵ

$$|| |\tilde{\phi}\rangle - |\phi'\rangle || < \epsilon.$$

Una idea para el problema anterior es descomponer el hamiltoniano H como suma de hamiltonianos que se pueden simular más simplemente, $H = H_1 + \cdots + H_k$, nos quedaría

$$|\phi(t)\rangle = e^{-i \sum_{i=1}^{k} H_i t} |\phi(0)\rangle .$$

No todos los H_i son conmutativos, por tanto no podemos separar la exponencial como producto

$$e^{-i \sum_{i=1}^{k} H_k t} \neq \prod_{i=1}^{k} e^{-iH_i t}.$$

Aunque la formula de Trotter nos indica el error de la factorización anterior

$$e^{-i \sum_{i=1}^{k} H_k t} = \prod_{i=1}^{k} e^{-iH_i t} O(t^2).$$

La fórmula anterior nos indica que para intantes pequeños de tiempo, o más general, para dos intantes de tiempo con diferencia, Δt, muy cercana a o podemos simular muy bien un hamiltoniano complejo. La contraposición es que a más pequeño el Δt más veces tendremos que recalcular el producto anterior para poder calcular instantes de tiempo mayores.

La descomposición del hamiltoniano puede ser extremadamente díficil dependiendo del caso, igualmente sabemos que podemos expresar un calquier operador como suma de ope-

radores de Pauli, aplicándolo al hamiltoniano:

$$H = \sum_{k_1,k_2,\ldots,k_n=1,x,y,z} a_{k_1,k_2\ldots,k_n}(\sigma_{k_1} \otimes \cdots \otimes \sigma_{k_n}).$$

Donde cada coeficiente es

$$a_{k_1,k_2\ldots,k_n} = \frac{1}{2^n}\mathrm{tr}(\sigma_{k_1} \otimes \cdots \otimes \sigma_{k_n} \otimes H).$$

5.2.5. Angular

Está codificación usa las puertas de rotación para codificar información, dado un vector $x = (x_1, \ldots, x_n)$ la codificación angular sería

$$|x\rangle = \bigotimes_{i=1}^{n} R(x_i)\,|0^n\rangle\,.$$

5.3. Redes neuronales cuánticas

Las redes neuronales son un tipo de modelo de aprendizaje automático inspiradas en el cerebro biológico, está compuesto por neuronas o perceptrones interconectados entre si y donde cada conexión tiene un peso asociado. Los pesos son optimizados durante el entrenamiento para cumplir un determinada tarea. El término red neuronal cuántica puede referirse a una red neuronal con sus neuronas implementadas en ordenadores cuánticos pero también puede referirse a arquitecturas construidas con circuitos variacionales cuánticos entrenables por propagación hacia atrás (backpropagation). Actualmente se discute si hay supremacia de las redes cuánticas sobre las clásicas.

En lo siguiente vamos a describir que es un circuito variacional y un perceptrón.

5.3.1. Circuitos variacionales

Son un esquema híbrido entre un uso de un commputador cuántica y uno clásico. Dado que todavía no hay ordenadores puramente cuánticos, estos circuitos se han vuelto muy populares para la implementación de algortimos cuánticos.

Un circuito está compuesto de una o varias puertas lógicas con parámetros ajustables, dada una entrada el circuito cuántico cálcula el coste de una función, a partir de este resultado actualizamos los parámetros en un ordenador clásico. El objetivo que tenemos en mente es el de minimizar o maximizar dicho coste. La figura 5.4 ilustra el funcionamiento de un circuito variacional.

Hay que destacar que en los circuitos variacionales al final realizamos una medición, por tanto los resultados son probabilísticos . Por supuesto lo podemos paralelizar pero aumentaría el número de qubits. A partir de estas dos mediciones actualizamos los parámetros, el motivo por el que no se sigue un análogo clásico de backpropagation basándonos en alguna minimización es porque es tremendamente costoso. De hecho si nos basasemos en el mínimo analítico, la forma de acceder a él suele involucrar calcular productos escales que son prácticamente irrealizables en computadores cuánticos. El producto escalar no es realizable por la propiedad de reversibilidad de las puertas lógicas, aunque hay métodos de inferirlo probabilísticamente.

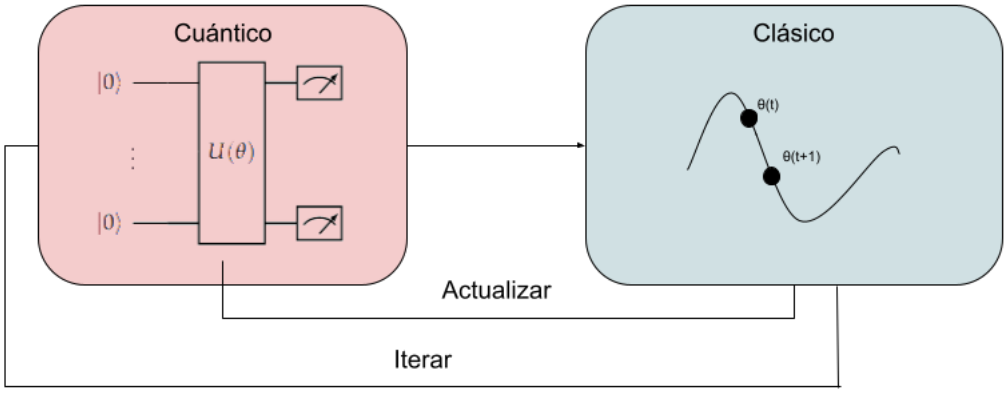


Figura 5.4.: Esquema de un circuito variacional.

5.3.2. Perceptrón

Si consideramos un conjunto de entradas $X = \mathbb{R}^d$, otro de salidas $Y = \mathbb{R}$, un $w \in \mathbb{R}^d$ y una función $f_w : X \to Y$ que dados unos datos da una predicción. Un perceptrón es una f_w no lineal

$$f_w(x) = \varphi(w \cdot x).$$

Deseamos encontrar un w tal que f_w prediga de forma correcta la mayor cantidad de elementos de X posibles. Por otro lado $\varphi : \mathbb{R} \to \mathbb{R}$ es una función no lineal y la denomonamos función de activación, la primera que se uso fue la escalonada

$$\varphi(a) = \begin{cases} 1\ ,\ a \geq 0, \\ 0\ ,\ a < 0. \end{cases}$$

Aunque también se ha optado tomar φ como la función sigmoide, la tengente hiperbólica o ReLU.

Entrenamos el perceptrón con la técnica del descenso del gradiente. Para cada (x_i, y_i) actualizamos los pesos de la siguiente manera

$$w^{(t+1)} = w^{(t)} + \eta[y_i - \varphi(w^{(y)} \cdot x_i)]x_i.$$

donde η es el ratio de aprendizaje. En la figura 5.5 tenemos el esquema de un perceptron.

Como hemos dicho, un perceptrón una funciones de activación no lineales. La no linealidad es un problema no trivial para la computación cuántica porque todas las operaciones cuánticas son lineales y porque los qubits tienen un carácter lineal. Por tanto en principio, no parece buena idea usar la computación cuántica para este tipo de aprendizaje automático, aunque la superposición y el entrelazamiento prometen buenas oportunidades. En lo siguiente vamos a explicar como poder realizar la función escalonada en un computador cuántico partiendo de la codificación en base y de la codificación en amplitud.

Primero desde codificación en base, vamos a suponer que nuestro conjunto de entrada $X = \{-1, 1\}^n$ y nuestra salida es $Y = \{-1, 1\}$. Para la codificación en bases partimos desde un conjunto de datos binarios, en este caso asociamos $-1 \to |0\rangle$ y $1 \to |1\rangle$.

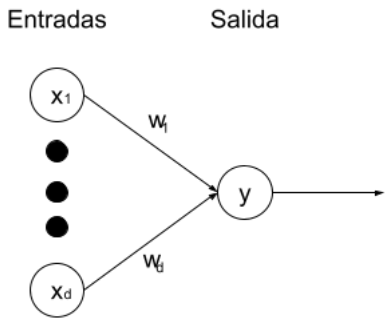


Figura 5.5.: Esquema de funcionamiento de un perceptron.

Definimos el siguiente estado cuántico que parte de la codificación de un elemento de X, n qubits que serán de registro, y tomamos t qubits extras

$$|\Psi_0\rangle = |x_1 x_2 \dots x_n\rangle \, |00\dots 0\rangle \, .$$

Con está codificación los pesos están codificados en operadores unitarios que actuán sobre los qubits de registro

$$U_w = U_n(w_n) \dots U_2(w_2) U_1(w_1) U_0.$$

Donde cada $U_i(w_i)$ actua sobre el registro i-ésimo, y son de la forma

$$U_i(w_i) = R_z(\frac{2\pi w_i}{n}) = \begin{pmatrix} e^{\frac{-2\pi i w_i}{2n}} & 0 \\ 0 & e^{\frac{2\pi i w_i}{2n}} \end{pmatrix} .$$

Con $U_0 = e^{i\pi} I$. Por tanto $U_0 \, |x_1 x_2 \dots x_n\rangle = e^{i\pi} \, |x_1 x_2 \dots x_n\rangle$, y el resto de puertas

$$x_i = -1 \; : \; U_i(w_i) \, |0\rangle = \begin{pmatrix} e^{\frac{-2\pi i w_i}{2n}} & 0 \\ 0 & e^{\frac{2\pi i w_i}{2n}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = e^{\frac{2\pi i w_i x_i}{2n}} \, |0\rangle$$

$$x_i = 1 \; : \; U_i(w_i) \, |1\rangle = \begin{pmatrix} e^{\frac{-2\pi i w_i}{2n}} & 0 \\ 0 & e^{\frac{2\pi i w_i}{2n}} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = e^{\frac{2\pi i w_i x_i}{2n}} \, |1\rangle \, .$$

Aplicar la puerta U_w a la codificación quedaría

$$U_w \, |x_1 x_2 \dots x_n\rangle = [e^{2\pi i (\frac{1}{2n} w \cdot x + \frac{1}{2})}] \, |x_1 x_2 \dots x_n\rangle = e^{2\pi i \theta} \, |x_1 x_2 \dots x_n\rangle \, .$$

Hemos sustituido $\theta = \frac{1}{2n} w \cdot x + \frac{1}{2}$ que sería la fase del vector propio $|x_1 x_2 \dots x_n\rangle$ del

operador U_w. Aplicando el algoritmo de estimación de fase cuántica 4.3.2 a la puerta U_w con la entrada $|\Psi\rangle$, obtendremos con cierta probabilidad el valor de θ con precisión t. A más qubits en usemos, más probable será obtener un mejor valor de *theta*, aunque no necesitamos dicho valor con exactitud.

El algoritmo de estamación de fase nos devolvería $|q_1 \dots q_t\rangle$ tales que

$$\frac{1}{2n} w \cdot x + \frac{1}{2} = \theta \approx q_1 \frac{1}{2^0} + q_2 \frac{1}{2^1} + \cdots + q_t \frac{1}{2^{t-1}}.$$

Como solo queremos saber si $w \cdot x \geq 0$, que equivale a que $\theta \geq \frac{1}{2}$, y solo depende del valor de q_1, principalmente, y del de q_2, en menor medida. A lo sumo, solo nos intersan 2 qubits de precisión consecuentemente podemos usar con $t = 2$ nos puede bastar, pero como dijimos cuanto más grande es el t con mejor probabilidad obtendremos el verdadero valor de q_1 y q_2.

De esta forma podemos implementar la función de activación φ y consecuentemente un perceptrón cuántico partiendo de la codificación en bases. La desventaja de este método es su ineficiencia con el número de qubits, la codificación en amplitud mitiga este defecto.

Veámos ahora como podemos implementarlo con la codificación en amplitud. Tomemos $x, w \in \{-1, 1\}^d$, si lo codificamos en amplitud necesitaremos $N \geq log(d)$ qubits, siendo esta

$$|\varphi_x\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} x_i |i\rangle \ , \ |\varphi_w\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} w_i |i\rangle .$$

Los pasos a seguir para implementar un perceptrón con está codificación son:

1. Preparar N registros a o: $|0\rangle^{\otimes N}$.
2. Aplicar U_x, para obtener la codificación de x: $|\varphi_x\rangle = U_x |0\rangle^{\otimes N}$.
3. Aplicar U_w, una puerta lógica que verifica $U_w |\varphi_w\rangle = |1\rangle^{\otimes N} = |d-1\rangle$.
4. Prepara un qubit extra, ancilla, $|0\rangle_a$.
5. Aplicar un *NOT* al bit ancilla, estando controlada por el resto de qubits.
6. Medimos la ancilla.

Tras el tercer paso tenemos el siguiente estado

$$|\varphi_{w,x}\rangle = U_w U_x |0\rangle^{\otimes N} = U_w |\varphi_x\rangle = \sum_i = 1^{d-1} c_i |i\rangle .$$

Donde el último coeficiente verfica

$$\frac{1}{d} w \cdot x = \langle \varphi_w | \varphi_x \rangle = \langle \varphi_w | U_w^\dagger U_w \varphi_x \rangle = \langle d-1 | U_w |\varphi_x\rangle \rangle = c_{d-1}.$$

Tras el quinto paso

$$CNOT_{N,a} |\varphi_{w,x}\rangle |0\rangle_a = \sum_i = 1^{d-2} c_i |i\rangle |0\rangle_a + c_{d-1} |d-1\rangle |1\rangle_a .$$

En el sexto medimos, la probabilidad de medir 1 es

$$P_a(1) = \frac{1}{d^2}|w \cdot x|^2.$$

Interpretamos lo siguiente: si el quibit ancilla está en el estado $|0\rangle$ el perceptrón está inactivo, y, si está en $|1\rangle$ se encuentra activo. Entonces la probabilidad de medir 1, representa la activación de la neurona.

Como comentario final, en el perceptrón cuántico a los pesos vamos a llamarlos parámetros. No hemos indicado como actualizar los pesos sino posibles implementaciones de un perceptrón. Si actualizamos los pesos podemos interpretar que un perceptrón es un tipo de circuito variacional.

5.3.3. Redes neuronales

La principal desventaja de un perceptrón residen en que aprende linealmente según los datos, por tanto para funciones no lineales no es la mejor opción. Esta desventaja se mitiga juntando varios perceptrones y conectándolos entre si, a esta conexión como ya dijimos antes es una red neuronal.

Los perceptrones de una red neuronal las podemos agrupar en capas, decimos que una red neuronal es prealimentada (red neuronal prealimentada) si podemos enumerar las capas de forma que la entradas de una capa son las salidas de la anterior y una capa posterior no sirve entrada a una anterior. Gráficamente podemos verlo en la figura 5.6.

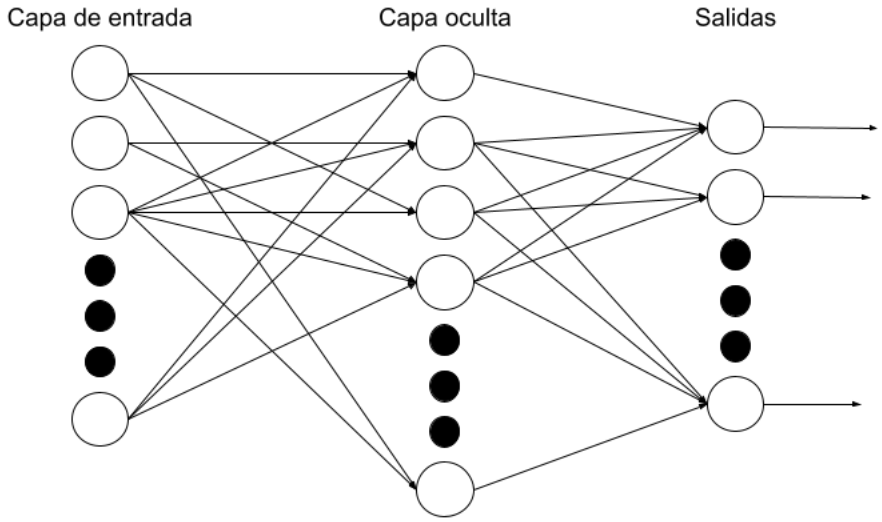


Figura 5.6.: Ilustración de una red neuronal prealimentada. Si todas las neuronas de una capa está conectadas con las de la siguiente lo decimos que están totalmente conectadas.

El modelo que aproximan las redes neuronales prealimentadas se puede escribir dadas $\varphi_1, \varphi_2, \ldots$ funciones de activación no lineales y matrices $W_1, W_2 \ldots$ de la dimensión adecuada

que representan pesos como

$$f_{W_1,W_2,\dots}(x) = \dots \varphi(W_2(\varphi(W_1 x)).$$

La función de perdida, $\mathcal{L}(W_1, W_2, \dots)$, es una función que indica el error que tenemos entre nuestra predicción y el objetivo.

La actualización de los pesos si la realizamos por descenso del gradiente es

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta \frac{\partial \mathcal{L}(W_1, W_2, \dots)}{\partial w_{ij}}.$$

Hasta ahora hemos dado generalidades de de redes neuronales prealimentadas, veamos como podemos implementar una red neuronal cuántica. Tenemos dos vías:

- Híbrida: después de cada capa medimos el resultado de cada perceptrón, procesamos los datos y se lo proporcionamos a la siguiente capa. Este método reduce la exigencia computacional cuántica, que es mas costosa. De esta forma algunos registros cuánticos podemos reutilizarlos.
- Coherente: implementamos toda la red en un hardware y solo medimos el resultado final. Este método requiere de más registros y por tanto es más demandante computacionalmente. La parte positva es el ahorro que hacemos en no tener que guardar los datos de forma clásica entre capas y luego tener que volver a codificarlos.

En general un red neuronal cuántica está dada en termino de sus perceptrones en cada capa. Como cada perceptron equivale a un puerta lógica $U(\theta_i)$ podemos decir que una red neuronal está dada por las puertás lógicas de cada capa. En estas redes los parámetros θ_i son los que queremos optimizar, serían lo análago a los pesos en redes neuronales clásicas. Tomaremos $U(\theta)$ la puerta lógica que corresponde a toda la red en el caso coherente, o, a una capa de perceptrones en el caso híbrido. Donde θ engloba a todos los parámetros.

Veamos un ejemplo de red neuronal, consideremos un problema de clasificación cuyos datos son $\{(x_i, y_i)\}_{i=1\dots N}$ con $y_i \in \{1, -1\}$ para cualquier $i = 1 \dots N$. Nuestra red constará de $n+1$ qubits, el útlimo qubit está preparado a 1, que implementa un circuito variacional descrito por el operador unitario $U(\theta)$ y guardamos el resultado en el qubit $n+1$-iesimo. La función de error que tomamos en este caso es

$$L(\theta) = -\sum_{i=1}^{N} y_i < x_i 1 | U^\dagger(\theta) \sigma_z^{n+1} >$$

donde σ_z^{n+1} aplica σ_z al qubit $n+1$ y la identidad a los demás. Multiples ejecuciones del circuito son necesarias para cálcular los valores esperados y poder usar el descenso del gradiente (o descenso del gradiente estocástico) en la minimización de la función de pérdida.

El entrenamiento de una red neuronal cuántica requiere de métodos de optimización clásicos y actualizaciones eficientes de los parámetros del circuito cuántico. La estructura general que implementa una red neuronal cuántica en terminos de un circuito variacional es la figura 5.7, donde una entrada $x \in \mathbb{R}^m$ la codificamos en n, $|\phi_x\rangle = U_F(x) |0\rangle^{\otimes n}$. Después el circuito variacional $U(\theta)$ realiza el proceso cuántico parámetrico que debe ser entrenado y la medición extrae el output.

La optimización de $U(\theta)$ se puede llevar a traves del método del descenso del gradiente, para ello tenemos que calcular las derivadas parciales respecto a los parámetros de la función

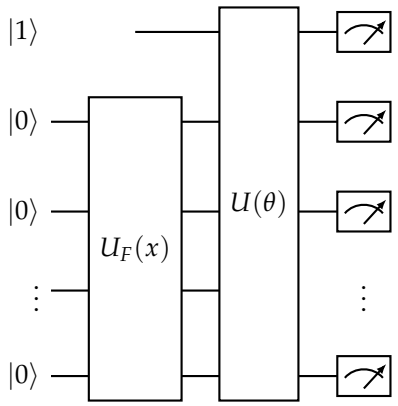


Figura 5.7.: Circuito variacional para el problema de clasificación.

de pérdida. En muchos casos la regla del cambio de parámetro (*parameter-shift rule*) es una estrategia interesante para calcular el gradiente. Veamos en detalle dicha regla.

Dado un operador A, en el caso del problema de clasificación que tenemos entre manos es σ_z^{n+1}, y el circuito cuántico $U(\theta)$, definimos la función paramétrica del modelo:

$$f(x,\theta) = \langle \phi_x | U^\dagger(\theta) A U(\theta) | \phi_x \rangle \,,$$

si podemos descomponer $U(\theta)$ como producto tensor de $U_i(\theta_i)$ puertas lógicas que solo afectan al qubit i-esimo. Entonces tenemos:

$$U(\theta) = V U_i(\theta_i) W$$

donde V y W son puertas parámetricas que depende del resto de parámetros. La derivada parcial de f es:

$$\begin{aligned}\frac{\partial f}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i} \langle \phi | U^\dagger(\theta_i) A U(\theta_i) | \phi \rangle \\ &= \langle \phi | \frac{\partial}{\partial \theta_i} U^\dagger(\theta_i) B U(\theta_i) | \phi \rangle + \langle \phi | U^\dagger(\theta_i) B \frac{\partial}{\partial \theta_i} U(\theta_i) | \phi \rangle\end{aligned}$$

donde $B = V^\dagger A V$ y $|\phi\rangle = W |\phi_x\rangle$. Un caso particular pero relevante es cuando la puerta $U_i(\theta_i)$ es de la forma

$$U_i(\theta_i) = e^{-i\theta_i H_i}$$

donde H_i es un operador autodajunto con dos valores propios distintos $\pm r$. Este requerimiento es verificado por cualquier puerta lógica aplicada a un qubit. Entonces la derivada parcial de f es:

$$\frac{\partial f}{\partial \theta_i} = \langle \phi' | B(-iH_i) | \phi' \rangle + \langle \phi' | (iH_i) B | \phi' \rangle \tag{5.1}$$

con $|\phi'\rangle = U_i(\theta_i) |\phi\rangle$. El siguiente lema es útil para calcular gradientes de circuitos variacionales.

Lema 5.1. *Sean B, C y D tres operadores lineales en un espacio de Hilbert. Entonces:*

$$C^\dagger BD + D^\dagger BC = \frac{1}{2}[(C+D)^\dagger B(C+D) - (C-D)^\dagger B(C-D)].$$

Demostración.

$$\begin{aligned}
&\frac{1}{2}[(C+D)^\dagger B(C+D) - (C-D)^\dagger B(C-D)] \\
&= \frac{1}{2}[(C^\dagger + D^\dagger)B(C+D) - (C^\dagger - D^\dagger)B(C-D)] \\
&= \frac{1}{2}[C^\dagger B(C+D) + D^\dagger B(C+D) - C^\dagger B(C-D) + D^\dagger B(C-D)] \\
&= \frac{1}{2}[C^\dagger BC + C^\dagger BD + D^\dagger BC + D^\dagger BD - C^\dagger BC + C^\dagger BD + D^\dagger BC - D^\dagger D] \\
&= \frac{1}{2}[2C^\dagger BD + 2D^\dagger BC] \\
&= C^\dagger BD + D^\dagger BC.
\end{aligned}$$

□

Aplicando este lema en (5.1) con $C = I$ y $B = -ir^{-1}H_i$

$$\frac{\partial f}{\partial \theta_i} = \frac{r}{2}(\langle \phi' | (I - ir^{-1}H_i)^\dagger B(I - ir^{-1}H_i)|\phi'\rangle - \langle \phi' | (I + ir^{-1}H_i)^\dagger B(I + ir^{-1}H_i)|\phi'\rangle). \quad (5.2)$$

Como H_i tiene dos valores propios $\pm r$, la serie de Taylor de $U_i(\theta_i)$ es de la siguiente forma:

$$U_i(\theta_i) = I\cos(r\theta_i) - ir^{-1}H_i\sin(r\theta_i),$$

como consecuencia

$$U_i(\frac{\pi}{4r}) = \frac{1}{\sqrt{2}}(I - ir^{-1}H_i).$$

Finalmente insertando la igualdad anterior en (5.2), obtenemos:

$$\begin{aligned}
\frac{\partial f}{\partial \theta_i} &= r(\langle \phi' | U_i^\dagger(\theta_i + \frac{\pi}{4r})BU_i(\theta_i + \frac{\pi}{4r})|\phi'\rangle - \langle \phi' | U_i^\dagger(\theta_i - \frac{\pi}{4r})BU_i(\theta_i - \frac{\pi}{4r})|\phi'\rangle) \\
&= r(f(\theta_i + \frac{\pi}{4r}) - f(\theta_i + \frac{\pi}{4r})).
\end{aligned}$$

La ecuación 5.3.3 se denomina regla de cambio de parámetro (*parameter-shift rule*) e implica que las derivadas parciales del circuito variacional pueden ser calculadas usando el propio circuito. Es decir usamos al propio circuito, $U(\theta)$, para calcular el modelo de la función y su gradiente. La actualización del parámetro se hace de forma clásica y la optimización se basa en un calculo cuántico del gradiente.

Desarrollando f,

$$f(x,\theta) = \langle \phi_x | U^\dagger(\theta)AU(\theta)|\phi_x\rangle = < |\phi_x\rangle, U^\dagger(\theta)AU(\theta)\,|\phi_x\rangle >,$$

vemos que involucra un producto escalar. En el apartado del perceptrón vimos como calcular el producto escalar entre dos elementos, usando alguno de los métodos vistos podemos

calcular $f(x, \theta)$. A partir de $f(x, \theta)$ podemos calcular su gradiente y tras obtener el gradiente podemos actualizar los parámetros por el método clásico descenso del gradiente (estocástico).

6. Series Temporales

Bibliografía

[CEMM98] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):339–354, January 1998.

[Hid19] Jack Hidary. *Quantum Computing: An Applied Approach*. 01 2019.

[Kay19] Alastair Kay. Quantikz, 2019.

[McM07] David McMahon. *Quantum Computing Explained*. 2007.

[Qis23] Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023.

[Sho97] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, October 1997.

[SP18] Maria Schuld and Francesco Petruccione. *Supervised Learning with Quantum Computers*. Springer Publishing Company, Incorporated, 1st edition, 2018.

[Sut19] Robert S Sutor. *Dancing with Qubits*. Packt Publishing, Birmingham, England, November 2019.

[ZJQ23] Amine Zeguendry, Zahi Jarir, and Mohamed Quafafou. Quantum machine learning: A review and case studies. *Entropy*, 25(2):287, February 2023.