QUANTUM CHEATSHEET

CÉDRIC HT

1. Gates

1.1. Pauli matrices.

The (2-dimensional) Pauli matrices are

$$I = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \qquad Y = \begin{pmatrix} & -i \\ i & \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}, \tag{1}$$

Because they are Hermitian, X and Y are unitarily diagonalizable as $X = H^+ZH$ and $Y = (HS^+)^+Z(HS^+)$. This immediately shows that the matrices X, Y, Z have two eigenvalues, ± 1. The corresponding *eigenstates* (normalized eigenvectors) are

$$\begin{aligned} \psi_{X,\pm} &= |\pm\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm 1 \end{pmatrix} & \psi_{Y,\pm} &= |\pm i\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm i \end{pmatrix} = S|\pm\rangle \\ \psi_{Z,+} &= |0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \psi_{Z,-} &= |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned} \tag{2}$$

A *Clifford matrix* is a unitary matrix M such that MAM^+ is a tensor product of Pauli matrices whenever A is too. They are important because quantum circuits made of Clifford gates can be simulated on classical computers in polynomial time (Gottesman-Knill theorem). Pauli matrices are Clifford.

1.2. Other basic gates.

Gate	Matrix form	Other forms	Bloch sph. rot.	Clifford?
H	$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	$\frac{X+Z}{\sqrt{2}}$	π around $\frac{\hat{\boldsymbol{x}}+\hat{\boldsymbol{z}}}{\sqrt{2}}$	Yes
$R_X(2\theta)$	$\begin{pmatrix} \cos(\theta) & -i\sin(\theta) \\ -i\sin(\theta) & \cos(\theta) \end{pmatrix}$	$\exp(-i\theta X)$	θ around $\hat{\boldsymbol{x}}$	No
$R_Y(2\theta)$	$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$	$\exp(-i\theta Y)$	θ around $\hat{\boldsymbol{y}}$	No
$R_Z(2\theta)$	$\begin{pmatrix} e^{-i\theta} & \\ & e^{i\theta} \end{pmatrix}$	$\exp(-i\theta Z)$	θ around $\hat{\boldsymbol{z}}$	No
$P(\theta)$	$\begin{pmatrix} 1 & \\ & e^{i\theta} \end{pmatrix}$			No
$\sqrt{X}$	$\frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}$		$\frac{\pi}{2}$ around $\hat{\boldsymbol{x}}$	Yes
S	$\begin{pmatrix} 1 & \\ & i \end{pmatrix}$	$P(\frac{\pi}{2}) = \sqrt{Z}$		Yes
T	$\begin{pmatrix} 1 & \\ & e^{i\frac{\pi}{4}} \end{pmatrix}$	$P(\frac{\pi}{4}) = \sqrt{S}$		No

The 2 qubits CNOT gate is a Clifford gate. In fact, the *Clifford group* (the group of all Clifford matrices of a given dimension, or equivalently, the normalizer of the Pauli group) is generated by $\{H, S, \text{CNOT}\}$ (adequatly tensored with the identity).

2. Measurements

2.1. Projective measurements.

What does it *mean* to measure a single qubit? First, one have to choose an orthonormal basis in which to perform the measurement. Consider a single qubit in state $|\psi\rangle$ and an orthonormal basis $\{|v_0\rangle, |v_1\rangle\}$. We can write

$$|\psi\rangle = \langle v_0|\psi\rangle|v_0\rangle + \langle v_1|\psi\rangle|v_1\rangle. \tag{3}$$

Then, measuring $|\psi\rangle$ in basis $\{|v_0\rangle, |v_1\rangle\}$ is an operation that gives 0 with probability $|\langle v_0|\psi\rangle|^2$ and 1 with probability $|\langle v_1|\psi\rangle|^2$. In other words, if $|\psi\rangle$ has components x_0 and x_1 in the basis $\{|v_0\rangle, |v_1\rangle\}$, then measuring in this basis gives 0 with probability $|x_0|^2$ and 1 with probability $|x_1|^2$.

It would be wrong to say that the system *is* in state $|v_0\rangle$ or $|v_1\rangle$ with probability $|x_0|^2$ or $|x_1|^2$. The system is in state $|\psi\rangle$, period. Probabilities are just about the *outcome of the measurement*.

2.2. **Observables and the Born rule.**

An *observable* is a simply a Hermitian matrix V. The *Born rule*, a postulate of quantum mechanics, states the following.

1. Physically speaking, the eigenvalues of V correspond to the different *quantized* values that a given property of a system can have (e.g. spin, energy). The eigenvectors of V represent the states of the system that have these definite values.
2. When *observing*, the system takes one of these definite values with a certain probability. Mathematically, of λ is an eigenvalue of V and $|\psi_\lambda\rangle$ the corresponding normalized eigenstate, then the probability of observing λ is $|\langle\psi_\lambda|\psi\rangle|^2$.

For example, in the 1 qubit case, V is unitarily diagonalizable (see Theorem 2.2.6) as

$$V = (\boldsymbol{v_0}, \boldsymbol{v_1}) \begin{pmatrix} \lambda_0 & \\ & \lambda_1 \end{pmatrix} \begin{pmatrix} \boldsymbol{v_0}^+ \\ \boldsymbol{v_1}^+ \end{pmatrix}, \tag{4}$$

where $\boldsymbol{v_0} = |v_0\rangle$ and $\boldsymbol{v_1} = |v_1\rangle$ are the normalized eigenvectors expressed as column vectors in the standard basis (also called computational basis), and where λ_0 and λ_1 are the corresponding eigenvalues. When performing the observation on a state $|\psi\rangle$, the probability of observing the value λ_0 is $|\langle v_0|\psi\rangle|^2$ and the probability of observing the value λ_1 is $|\langle v_1|\psi\rangle|^2$, assuming $\lambda_0 \neq \lambda_1$.

2.3. **Expected value of an observable.**

Given an observable V as in Equation 4, the expected value of the measurement is

$$\langle\psi|V|\psi\rangle = \lambda_0|\langle v_0|\psi\rangle|^2 + \lambda_1|\langle v_1|\psi\rangle|^2. \tag{5}$$

We also write $\langle V\rangle$ if the state $|\psi\rangle$ is clear from the context. Note that the unitary matrix $\begin{pmatrix} \boldsymbol{v_0}^+ \\ \boldsymbol{v_1}^+ \end{pmatrix}$ essentially rotates $|\psi\rangle$ so that observing V amounts to a measurement in the computational basis.

Given a state $|\psi\rangle = \psi_0|0\rangle + \psi_1|1\rangle$, the probabilities and expectation values of the Pauli matrices are:

Pauli A	$\mathbb{P}[1] = \left\|\langle\psi_{A,+}\|\psi\rangle\right\|^2$	$\mathbb{P}[-1] = \left\|\langle\psi_{A,-}\|\psi\rangle\right\|^2$	$\langle A\rangle$
X	$\frac{1}{2}\|\psi_0 + \psi_1\|^2$	$\frac{1}{2}\|\psi_0 - \psi_1\|^2$	$\frac{1}{2}\|\psi_0 + \psi_1\|^2 - \frac{1}{2}\|\psi_0 - \psi_1\|^2$

Pauli A	$\mathbb{P}[1] = \left\lvert\langle\psi_{A,+}\lvert\psi\rangle\right\rvert^2$	$\mathbb{P}[-1] = \left\lvert\langle\psi_{A,-}\lvert\psi\rangle\right\rvert^2$	$\langle A\rangle$
Y	$\frac{1}{2}\lvert\psi_0 + i\psi_1\rvert^2$	$\frac{1}{2}\lvert\psi_0 - i\psi_1\rvert^2$	$\frac{1}{2}\lvert\psi_0 + i\psi_1\rvert^2 - \frac{1}{2}\lvert\psi_0 - i\psi_1\rvert^2$
Z	$\lvert\psi_0\rvert^2$	$\lvert\psi_1\rvert^2$	$\lvert\psi_0\rvert^2 - \lvert\psi_1\rvert^2$

The same ideas transpose to the multi-qubit case. Common observables include "Pauli observables" on N qubits systems, which are linear combination[1] of terms of the form $A_0 \otimes A_1 \otimes \cdots \otimes A_N = A_0 A_1 \cdots A_N$, where the A_i's are Pauli matrices (including the identity). The observable is sparse if most of the A_i's are the identity matrix. For example, the following is a sparse Pauli observable on 5 qubits: $Z_1 - 2Z_5 = ZIIII - 2IIIIZ$ (I'm not sure what it represents though).

2.4. **Measurement operators.**

Projective measurements presented in Section 2.1 are a special case of the more general concept of quantum measurement, described in terms of *measurement operators*. Formally, a quantum measurement is a set $\{M_\lambda\}$ of matrices, called measurement operators, acting on the quantum system (i.e. of the right dimension), such that $\sum_\lambda M_\lambda^+ M_\lambda = I$ (this is called the *completeness equation*). The indices λ correspond to the different possible outcomes of the measurement. If the system is in state $|\psi\rangle$ and the measurement is applied, then outcome λ occurs with probability $\mathbb{P}[\lambda] = \langle\psi|M_\lambda^+ M_\lambda|\psi\rangle$. After observing λ, the state of the system becomes

$$\frac{M_\lambda|\psi\rangle}{\sqrt{\langle\psi|M_\lambda^+ M_\lambda|\psi\rangle}}. \tag{6}$$

A projective measurement is a quantum measurement where the measurement operator M_λ are the projector on the eigenspace of the eigenvalue λ. Taking V as in Equation 4, the measurement operators are $|v_0\rangle\langle v_0|$ and $|v_1\rangle\langle v_1|$.

Let's justify all that. First, note that $|\psi\rangle$, the operator $|\psi\rangle\langle\psi|$ is always Hemritian, and furthermore, $|\psi\rangle\langle\psi|^+|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle\langle\psi|$. Then, in the $\{|v_0\rangle, |v_1\rangle\}$ basis, $|v_0\rangle\langle v_0| = \begin{pmatrix} 1 & \\ & 0 \end{pmatrix}$ and $|v_1\rangle\langle v_1| = \begin{pmatrix} 0 & \\ & 1 \end{pmatrix}$. The completeness equation is thus satisfied. Finally, the probability of measuring λ_0 with a system in state $|\psi\rangle$ is $\mathbb{P}[\lambda_i] = \left\lvert\langle v_i|\psi\rangle\right\rvert^2$, as promised, and the state becomes

$$\frac{\langle v_i|\psi\rangle}{\left\lvert\langle v_i|\psi\rangle\right\rvert}|v_i\rangle \tag{7}$$

which is just $|v_i\rangle$ up to a global phase.

2.5. **POVM measurements.**

A *POVM* (Positive Operator-Valued Measure) is a just a mathematical re-expression of a quantum measurement as expressed in Section 2.4. Instead of measurement operators $\{M_\lambda\}$, we instead deal with a set of positive semidefinite operators $\{E_\lambda\}$ such that $\sum_\lambda E_\lambda = I$. Note that since E_λ is positive semidefinite, it make sense to talk about its square root $\sqrt{E_\lambda}$.

Measuring a system $|\psi\rangle$, the probability of outcome λ is $\mathbb{P}[\lambda] = \langle\psi|E_\lambda|\psi\rangle$, and the state of the system becomes

[1]The coefficients must be real numbers since the observable is a Hermitian matrix.

$$\frac{\sqrt{E_\lambda}|\psi\rangle}{\sqrt{\langle\psi|E_\lambda|\psi\rangle}} \tag{8}$$

If $\{M_\lambda\}$ is set of measurement operators in the sense of Section 2.4, and each M_λ is a projection matrix, then they form a POVM measurement. More generally, we can switch back and forth between the mathematical formalism of quantum measurements of Section 2.4 and POVM measurements by setting $M_\lambda = \sqrt{E_\lambda}$ and $E_\lambda = M_\lambda^+ M_\lambda$.[2]

2.6. **Application: measuring an ancilla qubit.**

Consider the state $|\psi\rangle = a_0|0\rangle|\psi_0\rangle + a_1|1\rangle|\psi_1\rangle$. Here, we think of the first qubit as an ancilla[3] qubit. Also, $|\psi_i\rangle$ need not be a quantum state, in the sense that $|\langle\psi_i|\psi_i\rangle|$ need not be equal to 1. What is the probability of observing the ancilla in the $|0\rangle$ state? What does it mean for the quantum system to just measure part of it?

Consider $P_0 = \begin{pmatrix} I & \\ & 0 \end{pmatrix}$ where I is the identity matrix on the state space of $|\psi_0\rangle$ and $|\psi_1\rangle$,[4] and $P_1 = \begin{pmatrix} 0 & \\ & I \end{pmatrix} = 1 - P_0$. The set $\{P_0, P_1\}$ forms a projective measurement. It is the *right* measurement in the sense that if *"event 0"* is observed, then the system will be in state $P_0|\psi\rangle = |0\rangle|\psi_0\rangle$ (modulo a normalization factor), which is consistent with the intuitive idea of *"having measured the ancilla in the* $|0\rangle$ *state"*. To repeat, *"event 0"* is *"having measured the ancilla in the* $|0\rangle$ *state"*. Applying Equation 8, the probability of this event occuring is thus

$$\mathbb{P}[0] = \langle\psi|P_0|\psi\rangle = |a_0|^2|\langle\psi_0|\psi_0\rangle|^2 = |a_0|^2\||\psi_0\rangle\|^2, \tag{9}$$

and the resulting state is

$$\frac{P_0|\psi\rangle}{\sqrt{\langle\psi|P_0|\psi\rangle}} = \frac{a_0}{|a_0|}\frac{1}{\||\psi_0\rangle\|}|0\rangle|\psi_0\rangle. \tag{10}$$

The term $\frac{a_0}{|a_0|}$ is just a global phase, and $\frac{1}{\||\psi_0\rangle\|}$ is a normalization factor.

2.7. **Application: generalized uncertainty principle.**

Consider a observable V and a system in state ρ. The variance of the measurement is $\sigma_V^2 = \langle(V - \langle V\rangle)^2\rangle$. For another observable W, some computation (involving Schwartz inequality) shows that

$$\sigma_V^2\sigma_W^2 \geq \left(\frac{1}{2i}\langle[V, W]\rangle\right)^2, \tag{11}$$

where $[V, W] = VW - WV$ is the *commutator* of V and W. The commutator measures how incompatible V and W are.

- If $[V, W] = 0$, then V and W commute, so they are simultaneously diagonalizable and share the same eigenvalues and eigenstates. Measuring V gives an outcome λ and projects the state of the system onto the

[2]By the way, a matrix A is positive semidefinite if and only if $A = B^+B$ for some matrix B.

[3]This is just a name for a qubit or a quantum register, and it has no deeper theoretical meaning.

[4]In other words, if $|\psi_i\rangle$ is a 2^n dimensional complex (non-necessarily unit) vector, then I is the identity on $\mathbb{C}^{2^n}$.

corresponding eigenspace. Measuring W next necessarily gives the same outcome. This somewhat corresponds to the fact that Equation 11 doesn't give any restriction on the simultaneous knowledge of V and W.

- If $[V, W] \neq 0$, then measuring V would project the system onto an eigenspace of V which might not fit neatly into an eigenspace of W. Thus there is some uncertainty about the value of W, and some information about the original state of the system has already been lost when mearuing V.

3. Density operators

3.1. **Definition.**

A *density operator* (or *density matrix*) is a matrix that describes the state of a quantum system. This is just a different mathematical formalism which uses a matrix unstead of vectors $|\psi\rangle$. A *density operator* is a positive semidefinite matrix ρ with trace 1. Given a vector representation of a quantum system $|\psi\rangle$, the associated density matrix is $\rho = |\psi\rangle\langle\psi|$. States of this form are called *pure* states. Note that a density matrix ρ represents a pure state if and only if it has rank 1.

Otherwise, ρ can be expressed as a convex combination of pure states $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, and is called a *mixed* state. Intuitively, this means that ρ has been prepared in state $|\psi_i\rangle$ with probability p_i.

Quantum physics then postulates the following.

1. ρ completely describes the physical state of the system, up to a global phase.[5]
2. Applying a unitary operator U on a system in state ρ gives a system in state $U\rho U^+$.
3. Given a system in state ρ and a set of measurement operators $\{M_\lambda\}$, outcome λ occurs with probability $\mathbb{P}[\lambda] = \operatorname{tr}(M_\lambda^+ M_\lambda \rho) = \operatorname{tr}(M_\lambda \rho M_\lambda^+)$, and the system is then in state
$$\frac{M_\lambda^+ \rho M_\lambda}{\operatorname{tr}(M_\lambda^+ M_\lambda \rho)}. \tag{12}$$
Note that if $\{M_\lambda\}$ is a projective measurement, i.e. each M_λ is a projector, then $\mathbb{P}[\lambda]$ is simply $\operatorname{tr}(M_\lambda \rho)$.
4. If we have a set of systems A_i with density operators ρ_i, then the composite system has density operator $\bigotimes_i \rho_i$.

If the system is in state ρ and V is an observable, then
$$\langle V \rangle = \operatorname{tr}(V\rho). \tag{13}$$
This is the correct definition in the following sense. Let $\{M_\lambda\}$ be the projections onto the eigenspaces of V, so that $V = \sum_\lambda M_\lambda$. Keep in mind that $M_\lambda^+ M_\lambda = M_\lambda$. Since $\langle V \rangle$ should be the expected measurement result, we have
$$\langle V \rangle = \sum_\lambda \lambda \mathbb{P}[\lambda] = \sum_\lambda \lambda \operatorname{tr}(M_\lambda^+ M_\lambda \rho) = \operatorname{tr}\left(\sum_\lambda \lambda M_\lambda \rho\right) = \operatorname{tr}(V\rho). \tag{14}$$

3.2. **Partial traces.**

[5] indeed, note that $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$ give the same density operator

Consider a compound system $\mathcal{B} \otimes \mathcal{A}$ in a state with density matrix $\rho^{\mathcal{B}\mathcal{A}}$. We'll see in Section 3.3 why the letters are reversed. Suppose we forget about $\mathcal{A}$. What is the state of $\mathcal{B}$ then? This is the question answered by the *partial trace*.

Suppose that $\rho^{\mathcal{B}\mathcal{A}} = \rho^{\mathcal{B}} \otimes \rho^{\mathcal{A}}$ for some pure or mixed state $\rho^{\mathcal{B}}$ of $\mathcal{B}$ and likewise for $\mathcal{A}$. Such a system is called *unentangled*. Then we define $\operatorname{tr}_{\mathcal{A}} \rho^{\mathcal{B}\mathcal{A}} := \rho^{\mathcal{B}}$, and likewise $\operatorname{tr}_{\mathcal{B}} \rho^{\mathcal{B}\mathcal{A}} := \rho^{\mathcal{A}}$. This is consistent with the idea that if $\mathcal{B}$ and $\mathcal{A}$ are independent systems, then discarding one does not change the state of the other.

In general however, $\rho^{\mathcal{B}\mathcal{A}}$ may not be a tensor product,[6] but $\operatorname{tr}_{\mathcal{B}}$ and $\operatorname{tr}_{\mathcal{A}}$ are still well defined by extending the special cases above by linearity. See also Equation 18 for an explicit formula when $\mathcal{B}$ and $\mathcal{A}$ are both 1 qubit systems.

Here is an interesting example. Suppose $\mathcal{B} \otimes \mathcal{A}$ is in state $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. Then $\operatorname{tr}_{\mathcal{B}} \rho^{\mathcal{B}\mathcal{A}} = \frac{I}{2}$, the completely mixed staten, a.k.a. the coin toss, which is classical, instead of the naively expected $|+\rangle\langle+|$. See also Equation 18.

3.3. **Endianness.**

Endianness is the way a bitstring should be read from some memory layout. A helpful slogan is *"In little-endian, little bits come first. In big-endian, big bits come first."* More formally, let's say that some memory has bits 1, 0, 1, 1 at addresses a, $a+1$, $a+2$ and $a+3$ respectively. In the *little-endian* convention, least significant bits have a lower address, and so the bitstring reads 1101. The *big-endian* convention is the opposite: the more significant bits are at lower addresses (i.e. they "come first"), and the memory above reads 1011.

Qiskit adopts the *little-endian* convention for qubits. For example, the input state state

$$\begin{array}{l} |1\rangle \text{ ——— } \cdots \\ |0\rangle \text{ ——— } \cdots \\ |1\rangle \text{ ——— } \cdots \\ |1\rangle \text{ ——— } \cdots \end{array}$$

reads $|1101\rangle = |13\rangle$, so that bits on the top are less significant. The endianness convention also impacts how matrices of multi-qubits unitary gates and tensor product of unitary gates are expressed. For example, the CNOT gate in little endian and big endian are respectively

$$\begin{array}{c} |00\rangle \\ |01\rangle \\ |10\rangle \\ |11\rangle \end{array} \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & & 1 \\ & & 1 & \end{pmatrix} \qquad \begin{array}{c} |00\rangle \\ |10\rangle \\ |01\rangle \\ |11\rangle \end{array} \begin{pmatrix} 1 & & & \\ & & & 1 \\ & & 1 & \\ & 1 & & \end{pmatrix} \tag{15}$$

From now on, we will use the little endian convention. If $\rho^{\mathcal{A}}$ and $\rho^{\mathcal{B}}$ are the density matrices of an unentangled compound system

$$\begin{array}{l} \rho^{\mathcal{A}} \text{ ——/—— } \cdots \\ \rho^{\mathcal{B}} \text{ ——/—— } \cdots \end{array}$$

then the density matrix of the whole system is

[6] In this case, we say that $\rho^{\mathcal{B}\mathcal{A}}$ is an *entangled* system.

$$\rho = \rho^{\mathcal{B}} \otimes \rho^{\mathcal{A}} = \begin{pmatrix} \rho^{\mathcal{B}}_{00}\rho^{\mathcal{A}} & \rho^{\mathcal{B}}_{01}\rho^{\mathcal{A}} & \cdots \\ \rho^{\mathcal{B}}_{10}\rho^{\mathcal{A}} & \rho^{\mathcal{B}}_{11}\rho^{\mathcal{A}} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{16}$$

One way to interpret this is that $\mathcal{A}$ is the "inner" system since its density matrix is "inside" ρ, whereas $\mathcal{B}$ is the "outer" system.

Now, consider the case where $\mathcal{A}$ and $\mathcal{B}$ are both 1 qubit systems, and the possibly entangled state

$$\rho = \begin{pmatrix} \rho_{00|00} & \rho_{00|01} & \rho_{00|10} & \rho_{00|11} \\ \rho_{01|00} & \rho_{01|01} & \rho_{01|10} & \rho_{01|11} \\ \rho_{10|00} & \rho_{10|01} & \rho_{10|10} & \rho_{10|11} \\ \rho_{11|00} & \rho_{11|01} & \rho_{11|10} & \rho_{11|11} \end{pmatrix} \tag{17}$$

Then, $\mathrm{tr}_{\mathcal{A}}$ is the "inner trace" whereas $\mathrm{tr}_{\mathcal{B}}$ is the "outer trace":

$$\begin{aligned} \mathrm{tr}_{\mathcal{A}}\,\rho &= \mathrm{tr}_{\mathcal{A}} \begin{pmatrix} \begin{pmatrix} \rho_{00|00} & \rho_{00|01} \\ \rho_{01|00} & \rho_{01|01} \end{pmatrix} & \begin{pmatrix} \rho_{00|10} & \rho_{00|11} \\ \rho_{01|10} & \rho_{01|11} \end{pmatrix} \\ \begin{pmatrix} \rho_{10|00} & \rho_{10|01} \\ \rho_{11|00} & \rho_{11|01} \end{pmatrix} & \begin{pmatrix} \rho_{10|10} & \rho_{10|11} \\ \rho_{11|10} & \rho_{11|11} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \mathrm{tr}\begin{pmatrix} \rho_{00|00} & \rho_{00|01} \\ \rho_{01|00} & \rho_{01|01} \end{pmatrix} & \mathrm{tr}\begin{pmatrix} \rho_{00|10} & \rho_{00|11} \\ \rho_{01|10} & \rho_{01|11} \end{pmatrix} \\ \mathrm{tr}\begin{pmatrix} \rho_{10|00} & \rho_{10|01} \\ \rho_{11|00} & \rho_{11|01} \end{pmatrix} & \mathrm{tr}\begin{pmatrix} \rho_{10|10} & \rho_{10|11} \\ \rho_{11|10} & \rho_{11|11} \end{pmatrix} \end{pmatrix} \\ \mathrm{tr}_{\mathcal{B}}\,\rho &= \mathrm{tr}_{\mathcal{B}} \begin{pmatrix} \begin{pmatrix} \rho_{00|00} & \rho_{00|01} \\ \rho_{01|00} & \rho_{01|01} \end{pmatrix} & \begin{pmatrix} \rho_{00|10} & \rho_{00|11} \\ \rho_{01|10} & \rho_{01|11} \end{pmatrix} \\ \begin{pmatrix} \rho_{10|00} & \rho_{10|01} \\ \rho_{11|00} & \rho_{11|01} \end{pmatrix} & \begin{pmatrix} \rho_{10|10} & \rho_{10|11} \\ \rho_{11|10} & \rho_{11|11} \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \rho_{00|00} & \rho_{00|01} \\ \rho_{01|00} & \rho_{01|01} \end{pmatrix} + \begin{pmatrix} \rho_{10|10} & \rho_{10|11} \\ \rho_{11|10} & \rho_{11|11} \end{pmatrix} \end{aligned} \tag{18}$$

4. The Bloch sphere

4.1. **Definition.**

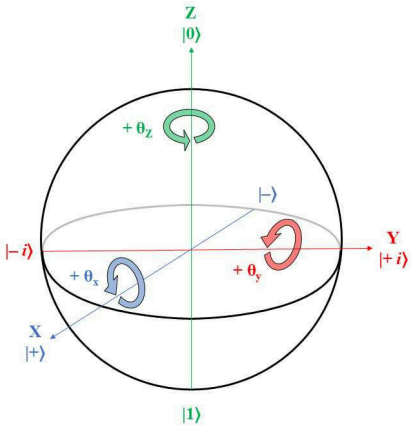
Consider a 1 qubit system. Its density matrix can always be decomposed as

$$\rho = \frac{1}{2}(I + v_x X + v_y Y + v_z Z). \tag{19}$$

where for $A \in \{X, Y, Z\}$, $v_a = \langle A \rangle$. The *Bloch vector* $\boldsymbol{v} = (v_x, v_y, v_z)$ is a point in the unit ball. If the state is pure, then $\|\boldsymbol{v}\| = 1$, so it is a point on the boundary, called the *Bloch sphere*. In particular:

- if $\rho = |0\rangle\langle 0|$ (resp. $|1\rangle\langle 1|$), then $\boldsymbol{v} = (0, 0, 1)$ (resp. $\boldsymbol{v} = (0, 0, -1)$), corresponding to the north (resp. south) pole;
- if $\rho = |\pm\rangle\langle \pm|$, then $\boldsymbol{v} = (\pm 1, 0, 0)$;
- if $\rho = |\pm i\rangle\langle \pm i|$, then $\boldsymbol{v} = (0, \pm 1, 0)$;

In the classical representation of the Bloch ball, the ambient 3D Euclidean space is oriented following the left hand rule. In particular, the eigenstates of a Pauli matrix $A \in \{X, Y, Z\}$ are on the axis $\hat{\boldsymbol{a}}$ of A, and the eigenstate with eigenvalue $\lambda \in \{-1, 1\}$ is at $\lambda\hat{\boldsymbol{a}}$.



If ρ is a mixed state, then $\|\boldsymbol{v}\| < 1$, so it is a point in the interior. The *maximally mixed state* is $\rho = \frac{I}{2}$ (i.e. $\boldsymbol{v} = 0$), corresponding to the center of the Bloch ball.

More generally, a quantum state ρ on n qubits can be decomposed as

$$\rho = \sum_A \langle A \rangle A = \frac{1}{2^n}\left(I + \sum_{A \neq I} \operatorname{tr}(A\rho) A\right) \tag{20}$$

where A ranges over all $4^n - 1$ non-identity tensor products of I's, X's, Y's and Z's. This represents a point in a $4^n - 1$ dimensional ball.

4.2. **Gate action on the Bloch sphere.**

Let's go back to a 1 qubit system. It just so happens that $A \in \{X, Y, Z\}$ acts on the the Bloch ball by a π rotation along the corresponding axis $\hat{\boldsymbol{a}}$. More generally, $R_A(\theta)$ acts by a (counterclockwise, or right hand) rotation of θ around $\hat{\boldsymbol{a}}$.

For example, consider the pure state $\rho = |0\rangle\langle 0|$ under the action of $R_X(\frac{\pi}{2}) = \begin{pmatrix} \cos(\theta/4) & -i\sin(\theta/4) \\ -i\sin(\theta/4) & \cos(\theta/4) \end{pmatrix}$. The new state $R_X(\frac{\pi}{2})\rho R_X(\frac{\pi}{2})^+$ corresponds to the pure state $R_X(\frac{\pi}{2})|0\rangle = \cos\frac{\pi}{4}|0\rangle - i\sin\frac{\pi}{4}|1\rangle$, which is indeed $|0\rangle$ rotated by $\frac{\pi}{4}$ around $\hat{\boldsymbol{x}}$.

Since $H = (X + Z)/\sqrt{2}$, it acts by a π rotation around the $\hat{\boldsymbol{x}} + \hat{\boldsymbol{z}}$ axis. More generally, $e^{i\theta H}$ acts by a θ rotation.

Since $\sqrt{X}\sqrt{X} = X$, we expect that X is a $\frac{\pi}{2}$ rotation around $\hat{\boldsymbol{x}}$. This is indeed the case, eventhough $R(\pi) \neq \sqrt{X}$. The reason is that $R(\pi)$ and $\sqrt{X}$ only differ by a global phase, which is not visible from the point of view of density matrices.

5. Hamiltonian shenanigans

5.1. **Hamiltonians.**

Let H be a Hamiltonian, which is just a fancy term for Hermitian matrix. Because it is Hermitian, it is unitarily diagonalizable and its eigenvalues (also called *energy levels*) are real.[7] The *ground energy* of H is its smallest eigenvalue, and a ground state is an eigenstate for that eigenvalue. Note that if $|\psi\rangle$ is a ground state, then $\langle H \rangle = \langle \psi | H | \psi \rangle$ is minimized.[8] A large class of problems can be reformulated

[7] This is in fact a necessary and sufficient condition.

[8] $\langle \psi | H | \psi \rangle = \lambda_{\text{ground}}$, but for non-ground states, and even non eigenstates, $\langle \varphi | H | \varphi \rangle = \sum_\lambda \langle v_\lambda | \varphi \rangle^2 \lambda \geq \sum_\lambda \langle v_\lambda | \varphi \rangle^2 \lambda_{\text{ground}} = \lambda_{\text{ground}}$, where $|v_\lambda\rangle$ are the eigenstates.

so that a solution is a ground state or highest energy state of a cleverly defined Hamiltonian. But finding these states is not easy…

5.2. **Quantum simulation.**

5.2.1. *Trotterization.*

There are two important formular to approximate the exponential of a sum of matrices.

- **Trotter's formula** (also known as the **Lie-Trotter formula** or even **Lie product formula**): if A and B are Hermitian operators (that need not commute), then

$$e^{-i(A+B)\theta} = e^{-iA\theta}\, e^{-iB\theta} + O(\theta^2). \tag{21}$$

This follows from a more general and exact formula:

$$e^{-i(A+B)\theta} = \lim_{n\to\infty} \left(e^{-iA\frac{\theta}{n}}\, e^{-iB\frac{\theta}{n}}\right)^n. \tag{22}$$

Iteratively applying it to a sum of N operators $H = \sum_j H_j$, we get:

$$e^{-iH\theta} = e^{-iH_1\theta} \cdots e^{-iH_N\theta} + O(N\theta^2). \tag{23}$$

In particular, Equation 21 can be "repeated" using the relationship $U(\theta) = U(\theta/k)^k$, where $U(\theta) = e^{-iH\theta}$ is the evolution operator. In other words, instead of taking one timestep θ, we can take k timesteps of θ/k:

$$e^{-i(A+B)\theta} = e^{-ik(A+B)\frac{\theta}{k}} = \left(e^{-iA\frac{\theta}{k}}\, e^{-iB\frac{\theta}{k}}\right)^k + O(\theta^2/k) \tag{24}$$

which, eventhough the theoretical error is still quadratic, is a better approximation in practice.

- **Suzuki's approximation** (also known as the **Suzuki-Trotter formula**): if A and B are Hermitian operators (that need not commute), then the first-order approximation with k timesteps is simply pulled from Equation 22:

$$e^{-i(A+B)\theta} = \left(e^{-iA\frac{\theta}{k}}\, e^{-iB\frac{\theta}{k}}\right)^k \tag{25}$$

More generally, if $H = \sum_j H_j$, we get:

$$e^{-iH\theta} = \left(e^{-iH_1\frac{\theta}{k}} \cdots e^{-iH_N\frac{\theta}{k}}\right)^k \tag{26}$$

The second-order approximation is

$$e^{-i(A+B)\theta} = e^{-iB\frac{\theta}{2}} e^{-iA\theta} e^{-iB\frac{\theta}{2}} + O(\theta^3). \tag{27}$$

Iteratively applying it to a sum of N operators $H = \sum_j H_j$, we get

$$e^{-iH\theta} = \left(e^{-iH_1\frac{\theta}{2}} \cdots e^{-iH_N\frac{\theta}{2}}\right)\left(e^{-iH_N\frac{\theta}{2}} \cdots e^{-iH_1\frac{\theta}{2}}\right) + O(\theta^3). \tag{28}$$

Write $S_2(\theta)$ for the right hand side of Equation 28 without the error term. The **$2n$-order Suzuki-Trotter approximation** is

$$S_{2n}(\theta) = S^2_{2(n-1)}(u_n\theta)\, S_{2(n-1)}((1-4u_n)\theta)\, S^2_{2(n-1)}(u_n\theta) \tag{29}$$

where

$$u_n = \frac{1}{4 - 4^{-(2n-1)}}. \tag{30}$$

As with Trotter's formula, Equation 27 can be decomposed into k timesteps to get a better approximation:

$$e^{-i(A+B)\theta} = e^{-ik(A+B)\frac{\theta}{k}} = \left(e^{-iB\frac{\theta}{2k}}\, e^{-iA\frac{\theta}{k}}\, e^{-iB\frac{\theta}{2k}}\right)^k + O(\theta^3/k). \tag{31}$$

- **Second-order Campbell-Baker-Hausdorff formula**: if A and B are Hermitian operators (that need not commute), then:

$$e^{-i(A+B)\theta} = e^{-iA\theta}\, e^{-iB\theta}\, e^{i\frac{1}{2}[A,B]\theta^2} + O(\theta^3). \tag{32}$$

If H is a Hamiltonian (possibly time-dependent), then applying any of the above formula to the evolution operator $U = e^{-iH\theta}$ is called the *trotterization* of the Hamiltonian.

5.2.2. *Time independent Hamiltonian.*

If a Hamiltonian H somehow acts on a quantum system, then the (time-independent) Schrödinger equation says that the state of the system evolves according to the flow defined by H:

$$i\hbar \frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = H|\psi(t)\rangle. \tag{33}$$

For an initial state $|\psi(0)\rangle$, the state of the system at time t is thus

$$|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle. \tag{34}$$

See Equation 37 for the time-dependent version.

Quantum simulation, also known as *Hamiltonian simulation* is the approximation of this flow by iteratively applying Equation 34 with small time increments δt. This requires to create a quantum circuit that approximates the unitary operator $e^{-iH\delta t}$, which is usually a hard problem. However, if the Hamiltonian is nice enough (e.g. a sparse Pauli operator), then the formula of Section 5.2.1 come in handy.

For example, the rotation gates $R_X(\theta) = \exp\left(-i\frac{\theta}{2}X\right)$ can be obtained by quantum simulation, where the Hamiltonian is X. Likewise for interation gates such as $R_{XY}(\theta)$.

It's important to note that the energy of the system $E(t) = \langle\psi(t)|H|\psi(t)\rangle$ does *not* change under the Hamiltonian flow.[9] First, if $|\psi(0)\rangle = |\psi_\lambda\rangle$ is an eigenstate with eigenvalue λ, then $|\psi(t)\rangle = e^{-i\lambda t}|\psi(0)\rangle$,[10] and so

$$\langle\psi(t)|H|\psi(t)\rangle = \langle e^{-i\lambda t}\psi(0)\big|H\big|e^{-i\lambda t}\psi(0)\rangle = e^{i\lambda t}e^{-i\lambda t}\langle\psi(0)|H|\psi(0)\rangle = \lambda, \tag{35}$$

so that the energy does not change. For a general $|\psi(0)\rangle = \sum_\lambda p_\lambda|\psi_\lambda\rangle$ we have $|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle = \sum_\lambda p_\lambda e^{-i\lambda t}|\psi_\lambda\rangle$, and

$$\langle\psi(t)|H|\psi(t)\rangle = \sum_\lambda |p_\lambda|^2\lambda = \langle\psi(0)|H|\psi(0)\rangle. \tag{36}$$

5.2.3. *Time dependent Hamiltonian.*

[9] However, the expectation value of some observable might.

[10] This follows by writing down the Taylor expansion of e^{-iHt}.

Consider a time-varying Hamiltonian $H(t)$. In this case, the Schrödinger equation reads

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = H(t)|\psi(t)\rangle \tag{37}$$

and the solution is $|\psi(t)\rangle = U(t)|\psi(0)\rangle$ where

$$U(t) = \exp\left(-i\int_0^t H(l)\,\mathrm{d}l\right). \tag{38}$$

up to a normalization constant so that $U(t)$ has determinant 1. By discretizing time, the integral can be approximated by $\sum_{n=0}^{N} H(n\delta t)\delta t$ for a small timestep δt. We can apply Trotter's formula (Equation 21) or Suzuki's approximation (Equation 27) to get

$$\begin{aligned} U(N\delta t) &\simeq e^{-iH(N\delta t)\delta t}\cdots e^{-iH(2\delta t)\delta t}\,e^{-iH(\delta t)\delta t} \\ &\simeq \left(e^{-iH(N\delta t)\frac{\delta t}{2}}\cdots e^{-iH(\delta t)\frac{\delta t}{2}}\right)\left(e^{-iH(\delta t)\frac{\delta t}{2}}\cdots e^{-iH(N\delta t)\frac{\delta t}{2}}\right). \end{aligned} \tag{39}$$

It is also possible to used repeated version (see Equation 24 and Equation 31).

6. Optimization

6.1. Quadratic Unconstrained Binary Optimization (QUBO) problems.

A QUBO problem is the minimization or maximization of a quadratic function of binary variables

$$\underset{x\in\{0,1\}^n}{\operatorname{argmax}}\, x^T Qx + c^T x \tag{40}$$

The cost function can equivalently be formulated as x^TQx since $x_i^2 = x_i$ for all i. One may also assume without loss of generality that Q is symmetric or even upper triangular.

Let's see how to translate Equation 40 into a quantum problem. Recall that if $|\psi\rangle$ is an eigenstate of H, then its energy is $\langle\psi|H|\psi\rangle$, and $H|\psi\rangle = \langle\psi|H|\psi\rangle|\psi\rangle$. The goal is to find a good Hamiltonian H such that if $|x\rangle$ is a pure state (so a binary string, modulo normalization), then (1) it is an eigenstate of H and (2) its energy is the value of the cost function, i.e. $\langle x|H|x\rangle = x^TQx + c^Tx$, or more conveniently, $H|x\rangle = (x^TQx + c^Tx)|x\rangle$.

Let's consider two examples:

1. Assume $Q = 0$ and c only have 1 at entry i entry. The objective function is then $x^TQx + c^Tx = x_i$. What operator H would satisfy $H|x\rangle = x_i|x\rangle$? Answer: $H = \frac{I-Z_i}{2}$, since
$$\frac{I-Z_i}{2}|x\rangle = \frac{1-(-1)^{x_i}}{2}|x\rangle \tag{41}$$
which is $0\times|x\rangle$ if $x_i = 0$ and $1\times|x\rangle$ if $x_i = 1$. The point is that $|x\rangle$ is always an eigenstate of H, and its eigenvalue is x_i.
2. If $c = 0$ and Q only has 1 at entry (i,j) for $i \neq j$, the same idea works with $H = \frac{I-Z_i}{2}\frac{I-Z_j}{2}$.

Finally, for a general Q and c, one considers the so-called *Ising Hamiltonian*[11]

$$H = \sum_{i,j} Q_{ij} \frac{I - Z_i}{2} \frac{I - Z_j}{2} + \sum_i c_i \frac{I - Z_i}{2}, \tag{42}$$

and by linearity, the eigenvalue of a bit-string $|x\rangle$ is $x^T Q x + c^T x$. Symbolically, we mapped a binary variable x_i with *values* $\{0, 1\}$ to an operator $\frac{I-Z_i}{2}$ with *eigenvalues* $\{1, -1\}$.

This mapping also works in reverse. Sometimes, the QUBO is already expressed in terms of *Ising variables*, i.e. variables in $\{1, -1\}$:

$$\operatorname*{argmax}_{z \in \{-1,1\}^n} z^T Q z + c^T z, \tag{43}$$

in which case the corresponding Ising Hamiltonian is even easier to write down

$$H = \sum_{i,j} Q_{ij} Z_i Z_j + \sum_i c_i Z_i. \tag{44}$$

Let's see why this works.

1. As before, consider the case where $Q = 0$ and c only has 1 at entry i. For a bit string $|s\rangle$, $s \in \{0, 1\}^n$, we have $H|s\rangle = (-1)^{s_i}|s\rangle$. So setting $z_i = 1 - 2s_i$ we get $H|s\rangle = z_i|s\rangle$, as expected.
2. If $c = 0$ and Q only has 1 at entry (i, j) with $i \neq j$, then $Z_i Z_j|s\rangle = (-1)^{s_i}(-1)^{s_j}|s\rangle = z_i z_j|s\rangle$.

So finally, by linearity again, a highest energy state $|s\rangle$ of H gives a solution $z = 1 - 2s$ of Equation 43.

Motto: A binary variable $x \in \{0, 1\}$ corresponds to an Ising variable $z \in \{1, -1\}$ (note the order).

6.2. **VQE.**

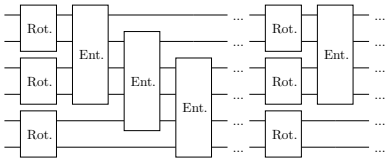
The *Variational Quantum Eigensolver* algorithm is an approach to approximate a ground state of a Hamiltonian. What's noteworthy is (1) the use of a classical optimizer (such as particle swarm optimization) and (2) the use of an *ansatz*, which is the choice of a subspace of the state space where we hope the ground state lies.

We start with a parameterized unitary operator $U(\boldsymbol{\alpha})$ which we call the *ansatz*.[12] This is just a fancy term for a highly-entangling circuit parametrized by a vector $\boldsymbol{\alpha}$ whose size if polynomial in the number of qubits. The parameters are usually used as angles in some rotation gates such as R_X or R_{YZ}. The ansatz then produces a state $|\psi(\boldsymbol{\alpha})\rangle := U(\boldsymbol{\alpha})|0\rangle$ and an energy $E(\boldsymbol{\alpha}) := \langle\psi(\boldsymbol{\alpha})|H|\psi(\boldsymbol{\alpha})\rangle$. The function E is considered to be a cost function that we minimize using classical means.

Qiskit provides helpful constructors for standard ansatzes such as `EfficientSU2` (implementation of SU(2)-spanned unitaries for low connectivity layouts), `ExcitationPreserving` (preserves the ratio of $|00\rangle$, $|01\rangle + |10\rangle$ and $|11\rangle$ states of the input superposition), or `QAOAAnsatz` (see Section 6). These are part of a wider family of circuits called *n-local circuits* composed of rotations and cascaded entanglements like

[11] An Ising Hamiltonian is usually presented with the form of Equation 44, but it is easy to see that Equation 42 is an equivalent formulation.

[12] Or *parametrized quantum circuit* (PQC) or *variational circuit*.



6.3. **Quantum Adiabatic Algorithm (QAA).**

Recall that in the time independent setting, the Hamiltonian flow does not change the energy of the system, see Section 5.2. This fact does not hold in the time dependent setting. However, the *adiabatic theorem* states that if the Hamiltonian $H(t)$ varies slowly enough, then a system that starts out in a ground state will remain in a ground state. In other words, and if $|\psi(0)\rangle$ is a ground state of $H(0)$, then $|\psi(t)\rangle$ is a ground state of $H(t)$.

QAA is a method that takes advantage of this fact. Assume that H_M is a time-independent Hamiltonian for which the ground state is known and easy to prepare, and that H_C is a Hamiltonian of interest whose ground state we want to find. We call these Hamiltonians the *mixer* and the *cost operator* respectively.

The idea is to start with H_M and slowly interpolate to H_C. Specifically, consider a small timestep δt and set $H(t) = \beta(t)H_M + \gamma(t)H_C$ with $\beta(t) = 1 - t/T$ and $\gamma(n) = t/T$, where $t \in [0, T]$ and T is the total evolution time. Let $\delta t = T/N$ for some $N \geq 1$ and consider

$$\begin{aligned} U(T) &= e^{-i\delta t H(N\delta t)} \dots e^{-i\delta t H(k\delta t)} \dots e^{-i\delta t H(0)} \\ &\simeq e^{-i\delta t H_C} \dots \left(e^{-i\delta t \beta(k\delta t) H_M}\, e^{-i\delta t \gamma(k\delta t) H_C}\right) \dots e^{-i\delta t H_M}. \end{aligned} \tag{45}$$

If N is large enough and if we start with a ground state $|\psi(0)\rangle$ if H_M, then $|\psi(T)\rangle = U(T)|\psi(0)\rangle$ will be close to a ground state of H_C.

It is important that H_M and H_C do not commute. If they do, then they share the same eigenspaces (i.e. energy levels) and $U(T)$ cannot change the energy of the system. In particular, $|\psi(T)\rangle$ will remain a ground state of H_M which might not be a ground state of H_C.

By replacing H_M by $-H_M$ and H_C by $-H_C$, this whole process transforms highest energy states into highest energy states.

6.4. **Quantum Approximate Optimization Algorithm (QAOA).**

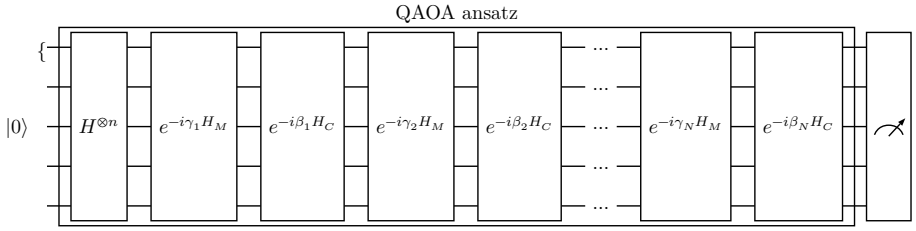
In QAA, the time-dependent Hamiltonian $H(t)$ is a linear interpolation between H_M and H_C. In QAOA, the interpolation is controlled by $2N$ variational parameters[13] $\beta_1, ..., \beta_N \in [0, \pi)$ and $\gamma_1, ..., \gamma_N \in [0, 2\pi)$. Starting with a highest energy state $|\psi(0)\rangle$ of the mixer H_M, we obtain a final state $|\psi(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle = U(\boldsymbol{\beta}, \boldsymbol{\gamma})|\psi(0)\rangle$, where

$$U(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \left(e^{-i\beta_N H_M}\, e^{-i\gamma_N H_C}\right) \dots \left(e^{-i\beta_1 H_M}\, e^{-i\gamma_1 H_C}\right). \tag{46}$$

The notations of Equation 45 and Equation 46 are a bit confusing, but roughly the idea is that β_k in Equation 46 plays the role of $\beta(k\delta t)\delta t$ in Equation 45, and likewise for γ.

A common mixer choice is $H_M = \sum_k X_k$ for which $H^{\otimes n}|0\rangle = |+\rangle^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{k=0}^{2^n-1} |k\rangle$ is a highest energy state. We then deal with the circuit

[13] i.e. parameters that shall be optimized classically



In Qiskit, this ansatz can be obtained by calling `QAOAAnsatz`.[14] By default, the mixer is $H_M = \sum_k X_k$ as above.

```
from qiskit.circuit.library import QAOAAnsatz
from qiskit.quantum_info import SparsePauliOp

Hc = SparsePauliOp(["ZZI", "ZIZ", "IZZ"])
N = 10  # Nb of checkpoints
qc = QAOAAnsatz(Hc, reps=N)
```

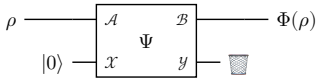
7. Quantum channels

A *quantum channel* (also called *quantum operation*) is just a mathematical representation of a (discrete-time) change or transformation[15] happening to a quantum system. This includes unitary operators, noise, and even measurements. Mathematically, a channel is a linear function $\Phi : M_{2^n}(\mathbb{C}) \longrightarrow M_{2^m}(\mathbb{C})$ that map density matrices to density matrices. For example, a unitary matrix U gives rise to a channel that maps ρ to $U\rho U^+$.[16] But not all channels arise in this way.

There are three main ways to mathematically represent channels:

- *Kraus representation.* A channel Φ can always be represented by a sequence of matrices $K_1, ..., K_N$ such that $\Phi(\rho) = \sum_i K_i \rho K_i^+$. The matrices must satisfy $\sum_i K_i^+ K_i = I$. The columns (resp. rows) of a K_i correspond to the classical states of the input (resp. output) system.

  Intuitively, this decomposition means that Φ maps the state ρ to $K_i \rho K_i^+$ with probability $\mathrm{tr}(K_i \rho K_i^+)$. So the K_i^+'s can be understood as measurement operators, see Section 3.1.
- *Steinspring representation.* The idea is to embed the system of interest $\mathcal{A}$ into a larger compound system $\mathcal{A} \otimes \mathcal{X}$ such that there is a unitary channel $\Psi : \mathcal{A} \otimes \mathcal{X} \to \mathcal{B} \otimes \mathcal{Y}$ that acts on $\mathcal{A}$ as desired. We then discard $\mathcal{Y}$. Intuitively, system $\mathcal{A}$ may not be closed, so we integrate the part of the environment it interacts with, namely $\mathcal{X}$.



- *Choi representation.* If the input (resp. output) system has n (resp m) classical states, then the Choi representation of Φ is a $nm \times nm$ matrix.

[14] Funny enough, in this context, QAOA is the acronym for "quantum alternating operator ansatz".

[15] i.e. the input and output systems need not be the same

[16] By the way, conjugating a positive semidefinite matrix by a (non-necessarily unitary) matrix always gives a positive semidefinite matrix.

$$J(\Phi) = \sum_{a,b} |a\rangle\langle b| \otimes \Phi(|a\rangle\langle b|)$$
$$= \begin{pmatrix} \Phi(|a_0\rangle\langle a_0|) & \cdots & \Phi(|a_0\rangle\langle a_{n-1}|) \\ \vdots & \ddots & \vdots \\ \Phi(|a_{n-1}\rangle\langle a_0|) & \cdots & \Phi(|a_{n-1}\rangle\langle a_{n-1}|) \end{pmatrix} \tag{47}$$

where a and b range over the classical states of X.

8. Noise

8.1. Types of noise.

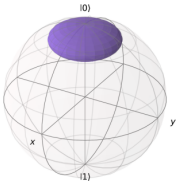
Coherent noise is kind of noise does not occur in the gates themselves. Formally, a ideal gate U is in fact physically implemented as a slightly different $\tilde{U}$.

On the other hand, *incoherent noise* is a kind of noise occurs "in the wires", i.e. when the qubit is (supposed to be) idle. They come in mainly three types:

- *Energy relaxation.* Whenever the physical implementation of a qubit maps the "low energy state" to $|0\rangle$ and the "excited state" to $|1\rangle$, it is expected that an exited qubit $|1\rangle$ will tend to decay to an unexcited qubit $|0\rangle$. So the wire maps $|0\rangle$ to $|0\rangle$ with probability 1 but $|1\rangle$ to $|0\rangle$ with some probability p. In Kraus form, an energy relaxation channel is given by

$$K_0 = \begin{pmatrix} 1 & \\ & \sqrt{1-p} \end{pmatrix}, \qquad K_1 = \sqrt{p}|0\rangle\langle 1| = \sqrt{p}\begin{pmatrix} & 1 \\ 0 & \end{pmatrix}. \tag{48}$$

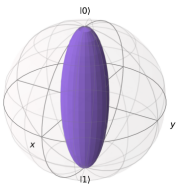
  Visually, this corresponds to squishing the Bloch ball onto the $|0\rangle$ pole.



- *Dephasing* occurs when a superposition of states becomes a classical probabilistic mixture, or in other words, quantum information becomes classical information. For example, the state $|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ tends to the completely mixed state. In density matrix form: $|+\rangle\langle+| = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ tends to become $\begin{pmatrix} 1/2 & \\ & 1/2 \end{pmatrix}$. In Kraus form, a dephasing channel is given by

$$K_0 = \sqrt{1-p}I, \qquad K_1 = \sqrt{p}Z. \tag{49}$$

  Visually, this corresponds to squishing the Bloch ball onto the Z axis, as indeed, states on this axis are classical mixtures.

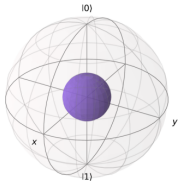


- *Depolarization* is when a (possible mixed) quantum state becomes the completely mixed state $\frac{1}{2}I$, which corresponds to the center of the Bloch

ball. A depolarization channel is given by $\mathcal{D}(\rho) = (1-p)\rho + p\,\mathrm{tr}(\rho)\frac{I}{2}$, or in Kraus form:

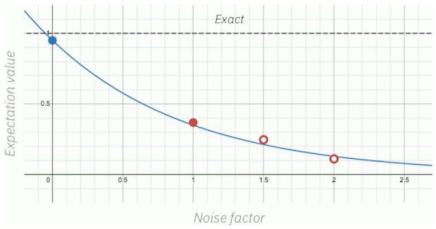
$$K_0 = \sqrt{1-p}I, \quad K_2 = \sqrt{\frac{p}{3}}X, \quad K_3 = \sqrt{\frac{p}{3}}Y, \quad K_4 = \sqrt{\frac{p}{3}}Z \tag{50}$$

Visually, this corresponds to squishing the Bloch ball onto the center, i.e. the completely mixed state.



8.2. **Terminology.**

- **Noise suppression**. Reduce or avoid the impact of noise. Usually happens before or during the execution, and usually required additional classical resources. For example, increasing the number of shots.
- **Noise mitigation**. Filter errors after they occur. In other words, we let errors happen and filter them out afterwards. Requires additional quantum resources.
- **Noise correction** (the crown jewel). Detect and fix errors as they occur during execution. Requires additional quantum and classical resources.
- **SPAM error**. Errors related to state preparation or measurement.
- **Cross-talk**. This is when a gate on one qubit affects another qubit, usually a neighboring idle qubit. This can be mitigated by using *dynamical decoupling*, where the idle qubit is "made busy" by passing through a circuit that ultimately puts it back to its current state, for example an even number of X gates.
- **ZNE (Zero-Noise Extrapolation)**. Only applicable to expectation value problems. The goal of ZNE is not to reduce noise but instead to progressively amplify. Then looking at the increasingly less reliable expectation values, we can extrapolate (with an exponential curve) to the noiseless case. Amplifying noise in such a controlled way is difficult, but the extrapolation part is actually even harder. This is due to the intrinsic instability of fitting an exponential curve to noisy samples whose scale is unknown (since we don't know the true ideal expectation value).



8.3. **Terminology on the IBM Quantum website.**

Calibration metrics of the IBM Quantum devices are reported using the following metrics.

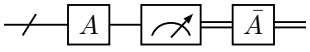
- $\boldsymbol{T_1}$ **(μs)**. Energy relaxation time in microseconds, i.e. the average time it takes for a qubit to decay from the excited $|1\rangle$ state to the ground state $|0\rangle$.
- $\boldsymbol{T_2}$ **(μs)**. Dephasing time in microseconds, i.e. the average time it takes for a qubit to go from $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ to a classical coin toss.
- **Frequency**. IBM quantum computers use superconducting qubits. The frequency is the energy difference between $|0\rangle$ and $|1\rangle$.
- **Readout assignment error**. the probability of a measurement returning the wrong value. This pertains to SPAM errors.
- **`Prob meas0 prep1`**. Probability of measuring 0 immediately after when preparing $|1\rangle$. This pertains to SPAM errors.
- **`Prob meas1 prep0`**. Similar.
- **Readout length (ns)**. The time it takes to perform a measurement, in nanoseconds.
- **ID error**. Error induced by having a qubit idle for a typical gate time. Can be though about the error in the physical implementation of the identity gate.
- $\boldsymbol{\sqrt{X}}$ **error**. Error induced by the $\sqrt{X}$ gate.
- **Single-qubit Pauli $\boldsymbol{X}$ error**. Similar.
- **CNOT error**. Similar. Important since it measures the error of applying an interaction between qubits.
- **Gate time (ns)**. The time it takes to perform a CNOT.

8.4. **Pauli Twirling.**

This is an error suppression technique aimed at coherent noise. The motivation is that coherent noise tends to accumulate quadratically, but after applying twirling, it only accumulates linearly. For example, if Λ is an incoherent noise channel, e.g. energy relaxation, then we can suppress it by twirling it into $\Lambda_A = A\Lambda A$, where A is a random Pauli matrix, and running the circuit multiple times (sampling a new Pauli matrix each time). However, in practice, we can't just twirl the noise channel in Λ because we can't isolate it from the rest of the circuit. Therefore, we cannot use Pauli twirling to suppress incoherent noise.

Instead, consider a noisy unitary gate $\tilde{U} = \Lambda U$, where Λ is some noise channel that we cannot isolate. By choosing a pair of Pauli matrices A and B such that $AU = UB$ (or in other words $AUB = U$), we have $A\tilde{U}B = A\Lambda AU$, and so we twirled the noise[17] without having direct access to it.

TREX (Twirled Readout Error eXtinction) is a readout error mitigation technique where instead of twirling a unitary gate, we twirl the measurement gate:



where $\bar{A}$ the matrix A applied to the readout. Since Pauli matrices are involutive, this is logically equivalent to a standard measurement as long as it is applied to expectation value problems, i.e. computing $\langle V \rangle$ for some observable V. This is why TREX cannot be applied to sampling.

[17] in an even more general way since A may not be equal to B.

A Trigonometric identities out the wazoo

$$\begin{cases} \cos(2x) = \cos^2(x) - \sin^2(x) \\ \qquad\quad = 2\cos^2(x) - 1 = 1 - 2\sin^2(x) \\ \sin(2x) = 2\sin(x)\cos(x) \\ \qquad\quad = (\cos(x) + \sin(x))^2 - 1 \end{cases} \qquad \begin{cases} \cos(\frac{x}{2}) = \operatorname{sgn}(\cos(\frac{x}{2}))\sqrt{\frac{1+\cos(x)}{2}} \\ \sin(\frac{x}{2}) = \operatorname{sgn}(\sin(\frac{x}{2}))\sqrt{\frac{1-\cos(x)}{2}} \end{cases}$$

$$\begin{cases} \cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y) \\ \sin(x \pm y) = \sin(x)\cos(y) \pm \cos(x)\sin(y) \end{cases} \qquad \begin{cases} \cos(x) + \cos(y) = 2\cos(\frac{x+y}{2})\cos(\frac{x-y}{2}) \\ \cos(x) - \cos(y) = -2\sin(\frac{x+y}{2})\sin(\frac{x-y}{2}) \\ \sin(x) \pm \sin(y) = 2\cos(\frac{x\pm y}{2})\cos(\frac{x\mp y}{2}) \end{cases}$$

$$\begin{cases} \cos(x)\cos(y) = \frac{\cos(x-y)+\cos(x+y)}{2} \\ \sin(x)\sin(y) = \frac{\cos(x-y)-\cos(x+y)}{2} \\ \sin(x)\cos(y) = \frac{\sin(x+y)+\sin(x-y)}{2} \end{cases} \qquad \begin{cases} \cos^2(x) = \frac{1+\cos(2x)}{2} \\ \sin^2(x) = \frac{1-\cos(2x)}{2} \end{cases}$$

The double-angle formula can be generalized using *Chebychev polynomials*: $\cos(nx) = T_n(\cos(x))$, where T_n, the Chebychev polynomial of the first kind, is defined by recurrence as

$$T_0(x) = 1, \qquad T_1(x) = x, \qquad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \tag{51}$$

Here are the first few: $T_2(x) = 2x^2 - 1$, $T_3(x) = 4x^3 - 3x$, $T_4(x) = 8x^4 - 8x^2 + 1$.

B Healthy linear algebra

B.1 **Diagonalizability.**

Let $A \in \mathrm{M}_n(\Bbbk)$. Its *characteristic polynomial* is $p_A(X) \coloneqq \det(XI_n - A)$ and its roots are the *eigenvalues* of A. The *minimal polynomial* m_A is the unique monic polynomial of least degree that vanishes on A.[18]

Theorem B.11.

1. *(Cayley-Hamilton) The matrix A is a root of its own characteristic polynomial.*
2. *p_A and m_A have exactly the same roots, which are the eigenvalues of A.*

If λ is an eigenvalue of A, then its multiplicity as a root of p_A is called its *algebraic multiplicity* of λ and denoted by $\operatorname{alg}(\lambda)$. The dimension of the corresponding eigenspace E_λ is called the *geometric multiplicity* and denoted by $\operatorname{geom}(\lambda)$.

Theorem B.12. $1 \leq \operatorname{geom}(\lambda) \leq \operatorname{alg}(\lambda) \leq n$.

Theorem B.13. *A is triangularizable iff p_A splits. In particular, every matrix is triangularizable over $\mathbb{C}$.*

If A is triangular, then its eigenvalues are on its diagonal. This means that the eigenspaces span the entire space, or in other words, there exists a basis of eigenvectors. Then, A is diagonalizable if and only if this basis is orthogonal.

If A and B commute, then they preserve each other's eigenspaces. Consequently, over $\mathbb{C}$, if two matrices commute, then they are simultaneously triangularizable, i.e. triangular in the same basis. Further, two diagonalizable matrices commute *if and only if* they are simultaneously diagonalizable.

[18] Recall that $\mathrm{M}_n(\Bbbk)$ is a PID.

Theorem B.14. *The following are equivalent:*

1. *A is diagonalizable;*
2. *$\Bbbk^n = \bigoplus_\lambda E_\lambda$ where λ ranges over the eigenvalues of A;*
3. *p_A splits and $\operatorname{geom}(\lambda) = \operatorname{alg}(\lambda)$ for all eigenvalue λ of A;*
4. *m_A only has simple roots.*

As we will see with SVD (Section B.6), a linear map $A : \Bbbk^n \to \Bbbk^n$ is essentially a rotation-stretching. Being diagonalizable means that, seen in the correct basis, A is just a stretching.

Theorem B.15. *(Jordan) If A is triangularizable then A is similar to a matrix of the form $\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & \end{pmatrix}$, also denoted $A_1 \oplus A_2 \oplus \cdots \oplus A_k$, called its* Jordan decomposition*, where*

1. *k is the number of eigenvalues of A;*
2. *A_i has size $\operatorname{alg}(\lambda_i) \times \operatorname{alg}(\lambda_i)$;*
3. *$A_i = \Lambda_{i,1} \oplus \cdots \oplus \Lambda_{i,\operatorname{geom}(\lambda_i)}$, where the $\Lambda_{i,j}$'s are called* Jordan blocks *and are of the form $\begin{pmatrix} \lambda_i & 1 & \\ & \ddots & \ddots \\ & & \end{pmatrix}$*
4. *the size of the biggest Jordan block for λ_i corresponds to its multiplicity as a root of m_A.*

Lastly, this decomposition is unique up to permutation of the Jordan blocks.

As a consequence, if A is triangularizable, then $\operatorname{tr} A = \sum_\lambda \operatorname{alg}(\lambda)\lambda$ and $\det A = \prod_\lambda \lambda^{\operatorname{alg}(\lambda)}$, where λ ranges over the eigenvalues of A.

B.2 Factorization.

A complex matrix $A \in \mathrm{M}_n(\mathbb{C})$ is *normal* if $A^+A = AA^+$. It is *Hermitian* (symmetric in the real case) if $A = A^+$. It is *unitary* (orthogonal in the real case) if $AA^+ = I_n$, or equivalently, if the columns (or rows) of A form an orthonormal basis.

Theorem B.26. *(Spectral theorem)*

1. *The following are equivalent: (1) A is normal and (2) A is unitarily diagonalizable.*
2. *The following are equivalent: (1) A is Hermitian; (2) A is* self-adjoint*, i.e. $(A\boldsymbol{x})^+\boldsymbol{y} = x^+(A\boldsymbol{y})$, or in braket notation, $\langle Ax|y\rangle = \langle x|Ay\rangle$[19]; and (3) A is unitarily diagonalizable and its eigenvalues are real*
3. *The following are equivalent: (1) A is unitary; (2) A preserves inner products, i.e. $(A\boldsymbol{x})^+(A\boldsymbol{y}) = \boldsymbol{x}^+\boldsymbol{y}$, or in braket notation, $\langle Ax|Ay\rangle = \langle x|y\rangle$; and (3) A is a normal matrix and its eigenvalues lie in the unit circle.*

Spectral decomposition is a method that decomposes a normal matrix A into $A = B - C$, and where B and C have orthogonal support, and where the eigenvalues of B (resp. C) are ≥ 0 (resp. > 0). First, recall that since A is unitarily diagonalizable, its eigenspaces are pairwise orthogonal. Write $P_{\geq 0}$ and $P_{<0}$ for the projectors onto $\bigoplus_{\lambda \geq 0} E_\lambda$ and $\bigoplus_{\lambda < 0} E_\lambda$ respectively, where λ ranges over the eigenvalues of A. We can just set $B = AP_{\geq 0}$ and $C = -AP_{<0}$. Since A is normal, it is diagonalizable, and so are B and C, all in the same basis.

[19] at which point the notation $\langle x|A|y\rangle$ becomes well-defined

B.3 **Positive semidefinite matrices.**

A *positive semidefinite* matrix $A \in M_n(\mathbb{C})$ (also called *non-negative* matrix, also denoted by $A \geq 0$) is one that satisfies $\boldsymbol{x}^+ A\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \mathbf{0}$. In braket notation, this reads $\langle x|Ax\rangle \geq 0$. Geometrically, that means that the angle between $A\boldsymbol{x}$ and $\boldsymbol{x}$ is $\leq \frac{\pi}{2}$. There are a few equivalent characterizations:

- A is Hermitian and all its eigenvalues are ≥ 0;
- $A = B^+B$ for some matrix B.

B.4 **Square roots.**

A matrix A has a *square root* if there exists a matrix B such that $B^+B = A$. If A is positive-semidefinite, then:

- it always has a square root;
- any two square roots $A = B^+B = C^+C$ are related by rotation, i.e. $B = CU$ for some unitary matrix U,
- there exists a unique matrix B such that $BB = A$.

B.5 **Matrix exponential.**

The exponential of a matrix A is defined naturally in terms Taylor series: $e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}$. This series always converges. Here are a few facts.

- If A and B commute, then $e^{A+B} = e^A e^B$.
- If A is involutive, then $e^{i\theta A} = \cos(\theta)I + i\sin(\theta)A$.
- An operator $U = e^{iA}$ is unitary if and only if A is Hermitian. In this case, U is called the *evolution operator* of A.
- If B is invertible, then $e^{BAB^{-1}} = Be^A B^{-1}$.

B.6 **Singular value decomposition.**

Consider a matrix $A \in \mathrm{M}_{m\times n}(\mathbb{C})$. A *singular decomposition* of A is a product $A = UDV^+$ where $U \in \mathrm{M}_m(\mathbb{C})$ and $V \in \mathrm{M}_n(\mathbb{C})$ are unitary, and $D \in \mathrm{M}_{m\times n}(\mathbb{C})$ is diagonal. The diagonal entries $s_1, ..., s_{\min(m,n)}$ of D are the *singular values* of A. Note that $AV_{-,i} = \sigma_i U_{-,i}$ for $i \leq \min(m, n)$.

Here's what that means. The matrix A forms a linear map $\mathbb{C}^n \to \mathbb{C}^m$, and the columns of V (resp U) form an orthonormal basis of $\mathbb{C}^n$ (resp. $\mathbb{C}^m$). Then, A maps the columns of V to the columns of U, modulo some stretching given by the singular values. So on the one hand, mapping orthonormal basis to orthonormal basis is essentially a rotation, and on the other hand, multiplying by a diagonal matrix is a stretching. SVD illustrates that every (at least finite dimensional) linear map is just a rotation-stretching.

RIKEN Quantum, RIKEN, Japan
Email address: cedric.hothanh@riken.jp