Quantum Algorithms
Lecture 18
Quantum probability II

Zhejiang University

Distance functions for density matrices

Defining the distance

In practice, various mixed states are always specified with some precision, so we need to somehow measure “distance” between density matrices. What would the most natural definition of this distance be? To begin with, let us ask the same question for probability distributions.

Faulty device

Let w be the probability distribution of an outcome produced by some device. Suppose that the device is faulty, i.e., with some probability ε it goes completely wrong, but with probability $1 - \varepsilon$ it works as expected. What can one tell about the actual probability distribution w' of the outcome of such a device? The answer is

$$\sum_j \left| w_j' - w_j \right| \leq 2\varepsilon$$

Probability distribution

$$\sum_j |w'_j - w_j| \le 2\varepsilon$$

Conversely, if the given inequality is true, we can represent w' as the probability distribution produced by a pipeline of two processes: the first generates j according to the distribution w, whereas the second alters j with total probability $\le \varepsilon$.

L1 norm

L1 Norm is the sum of the magnitudes of the vectors in a space. It is the most natural way to measure distance between vectors, that is the sum of absolute difference of the components of the vectors. In this norm, all the components of the vector are weighted equally.

Example:

$$X = \binom{3}{i}$$

$$\|X\|_1 = |3| + |i| = 3 + 1 = 4$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad |\mathbf{x}|_1 = \sum_{r=1}^{n} |x_r|$$

L1 norm

We conclude that the natural distance between probability distributions is given by the l^1 norm

$$\|w - w'\|_1 = \sum_j \left| w_j' - w_j \right|$$

Now we will generalize this definition to arbitrary density matrices.

Trace norm

The trace norm of an operator $A \in L(N)$ is

$$\|A\|_{tr} = Tr(\sqrt{A^\dagger A})$$

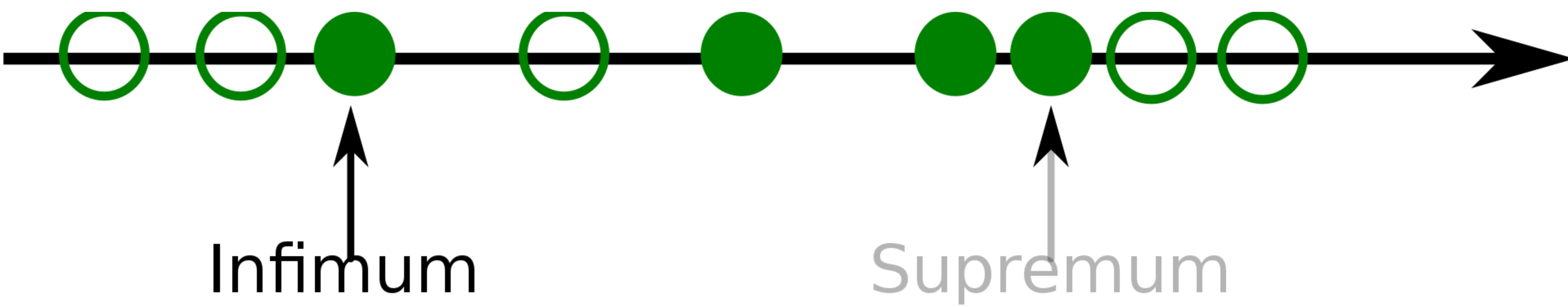
For Hermitian operators, the trace norm is the sum of the moduli of the eigenvalues.

Additional explanation:

For unitary matrix U, $U^\dagger U = I$, trace is equal to sum of diagonal elements, therefore, $\|U\|_{tr} = N$, where N is dimension of the matrix (number of rows/columns)

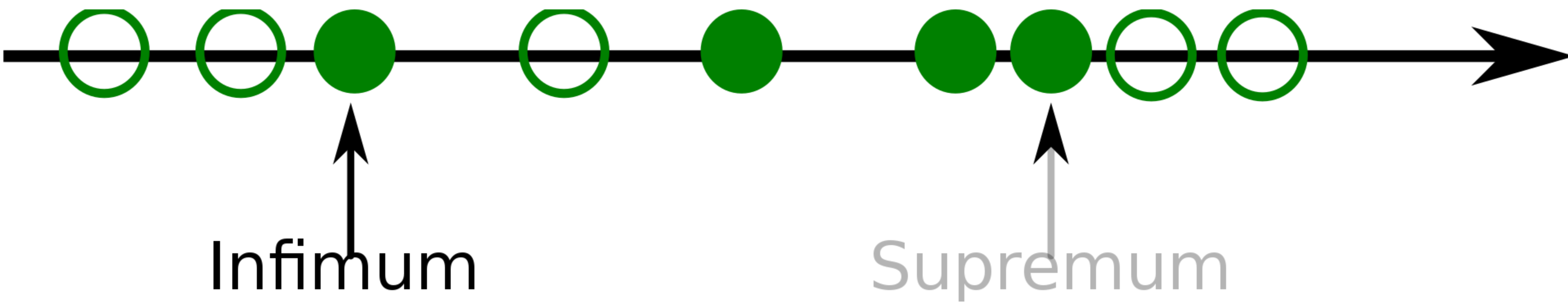
Norm equations

$$\|A\|_{\mathrm{tr}} = \sup_{B \neq 0} \frac{|\operatorname{Tr} AB|}{\|B\|} = \max_{U \in \mathbf{U}(\mathcal{N})} |\operatorname{Tr} AU|$$



Norm equations

$$\|A\|_{\mathrm{tr}} = \inf\left\{ \sum_k \big\| |\xi_k\rangle \big\| \, \big\| |\eta_k\rangle \big\| \; : \; \sum_k |\xi_k\rangle\langle\eta_k| = A \right\}$$



Properties of the trace norm

a) $\|AB\|_{tr}, \|BA\|_{tr} \leq \|A\|_{tr}\|B\|$
b) $|Tr\ A| \leq \|A\|_{tr}$
c) $\|Tr_M A\|_{tr} \leq \|A\|_{tr}$
d) $\|A \otimes B\|_{tr} \leq \|A\|_{tr}\|B\|_{tr}$

Trace norm vs L1 norm

The following lemma shows why the trace norm for density matrices can be regarded as the analogue of the l^1 -norm for probability distributions.

Lemma

Let $N = \bigotimes_j N_j$ be a decomposition of N into the direct sum of mutually orthogonal subspaces. Then for any pair of density matrices ρ and γ,

$$\sum_j \left| P(\rho, N_j) - P(\gamma, N_j) \right| \leq \|\rho - \gamma\|_{tr}$$

Proof of lemma

$$\sum_j \left|P(\rho, N_j) - P(\gamma, N_j)\right| \leq \|\rho - \gamma\|_{tr}$$

The left-hand side of the inequality can be represented in the form $Tr((\rho - \gamma)U)$, where $U = \sum_j (\pm \Pi_{N_j})$. It is clear that U is unitary. We then apply the representation of the trace norm in the form:

$$\|A\|_{\mathrm{tr}} = \sup_{B \neq 0} \frac{|\operatorname{Tr} AB|}{\|B\|} = \max_{U \in \mathbf{U}(\mathcal{N})} |\operatorname{Tr} AU|$$

Fidelity distance

There is another commonly used distance function on density matrices, called the fidelity distance. Let $\rho, \gamma \in L(N)$. Consider all possible purifications of ρ and γ over an auxiliary space F of dimension dim $F = \dim N$; these are pure states $|\xi\rangle, |\eta\rangle \in N \otimes F$. Then the fidelity distance between ρ and γ is

$$d_F(\rho, \gamma) \stackrel{\text{def}}{=} \min\Big\{ \big\| |\xi\rangle - |\eta\rangle \big\| \ : \ \mathrm{Tr}_{\mathcal{F}}(|\xi\rangle\langle\xi|) = \rho, \ \mathrm{Tr}_{\mathcal{F}}(|\eta\rangle\langle\eta|) = \gamma \Big\}$$

Fidelity distance

$$d_F(\rho, \gamma) \stackrel{\text{def}}{=} \min\Big\{ \big\| |\xi\rangle - |\eta\rangle \big\| \;:\; \operatorname{Tr}_{\mathcal{F}}(|\xi\rangle\langle\xi|) = \rho,\; \operatorname{Tr}_{\mathcal{F}}(|\eta\rangle\langle\eta|) = \gamma \Big\}$$

It is related to a quantity called fidelity:

$$F(\rho, \gamma) \stackrel{\text{def}}{=} \max\Big\{ \big|\langle\xi|\eta\rangle\big|^2 \;:\; \operatorname{Tr}_{\mathcal{F}}(|\xi\rangle\langle\xi|) = \rho,\; \operatorname{Tr}_{\mathcal{F}}(|\eta\rangle\langle\eta|) = \gamma \Big\}$$

Remark about purifications

One can show that the condition $dimF = dimN$ in these definitions can be relaxed: it is sufficient to require that $dimF \geq \max\{rank(\rho), rank(\gamma)\}$. Thus, any auxiliary space F will do, as long as it allows purifications of ρ and γ.

Fidelity properties

a) $d_F(\rho,\gamma) = \sqrt{2\left(1 - \sqrt{F(\rho,\gamma)}\right)}$;

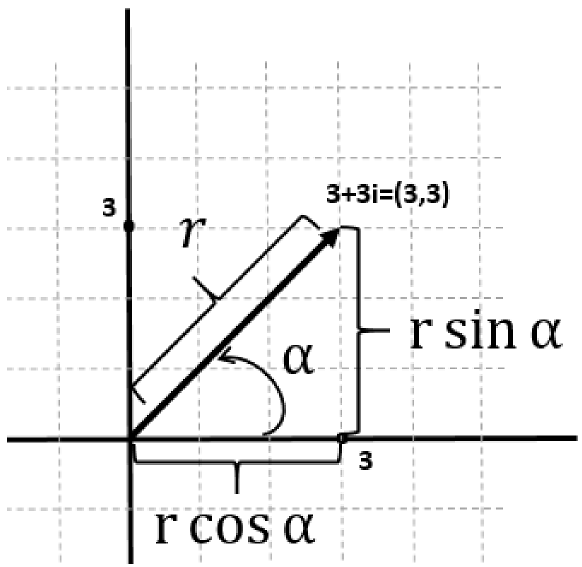
b) $F(\rho,\gamma) = \left\|\sqrt{\rho}\sqrt{\gamma}\right\|_{\mathrm{tr}}^2$;

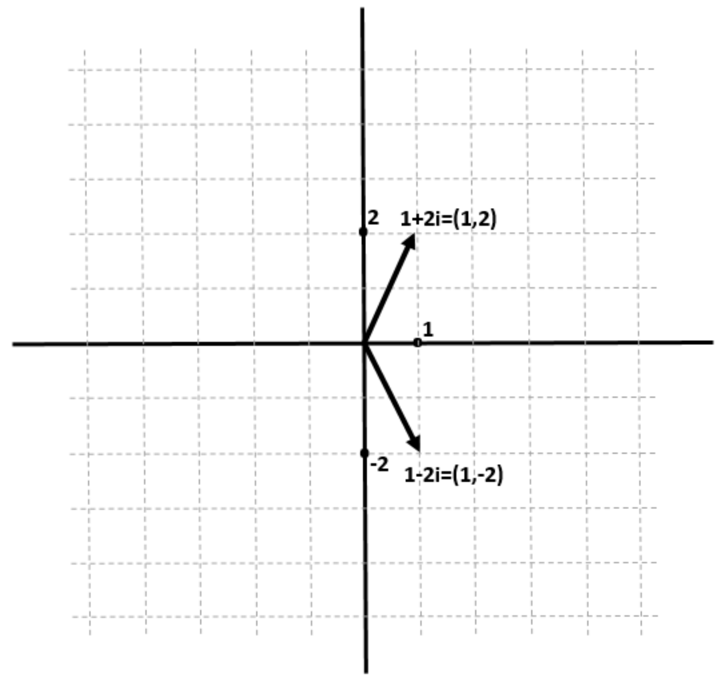
c) $\left(1 - \frac{\|\rho-\gamma\|_{\mathrm{tr}}}{2}\right)^2 \le F(\rho,\gamma) \le 1 - \left(\frac{\|\rho-\gamma\|_{\mathrm{tr}}}{2}\right)^2.$

Bloch sphere

Complex numbers

$$r(cos\alpha + i \cdot sin\alpha) = r \cdot e^{i\alpha}$$



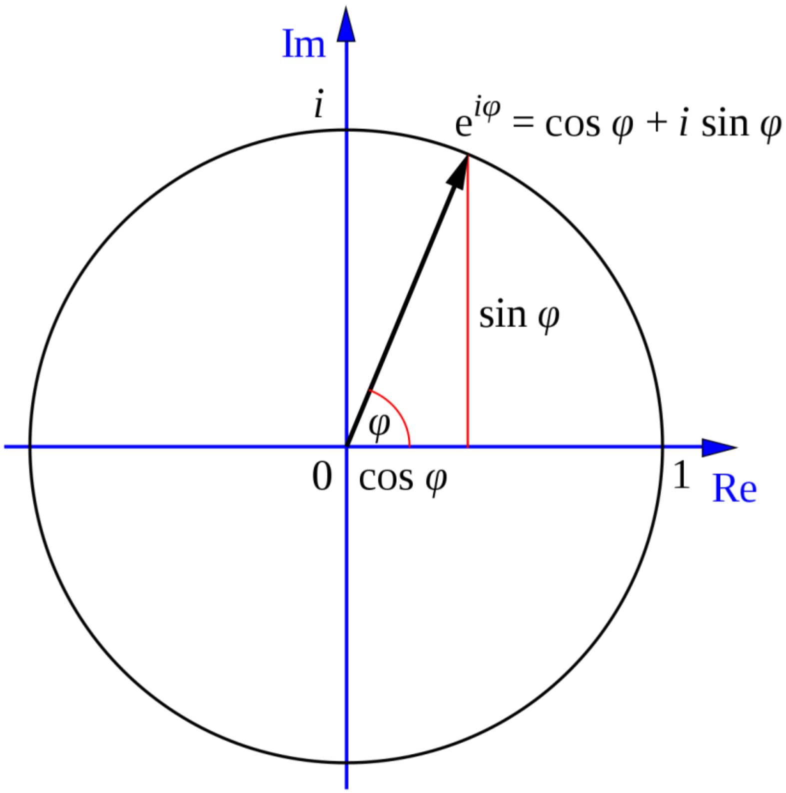


Complex numbers

In quantum computing we usually deal with complex numbers with absolute value 1.

$$e^{i\varphi}$$

$$e^{i\pi} + 1 = 0$$



State of a qubit

We have $|\psi\rangle = \mathrm{a}|0\rangle + \mathrm{b}|1\rangle$, where a and b are complex numbers.

There is alternative way to represent a state:

$$|\psi\rangle = e^{i\delta}(\cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle)$$

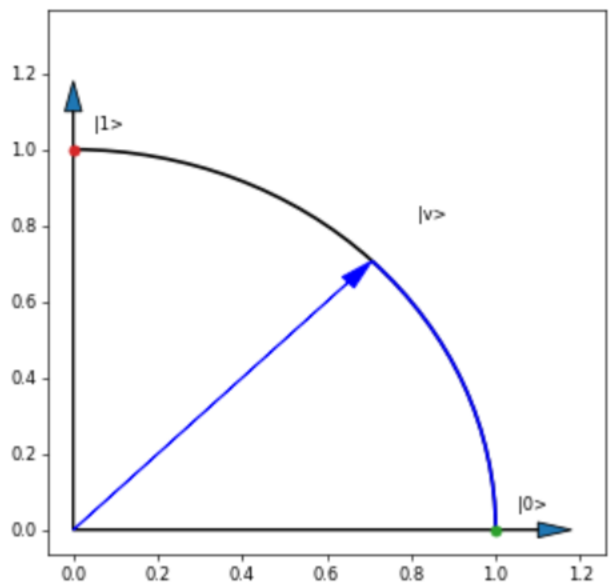
Here, $e^{i\delta}$ is a global phase, so it can be discarded. We get

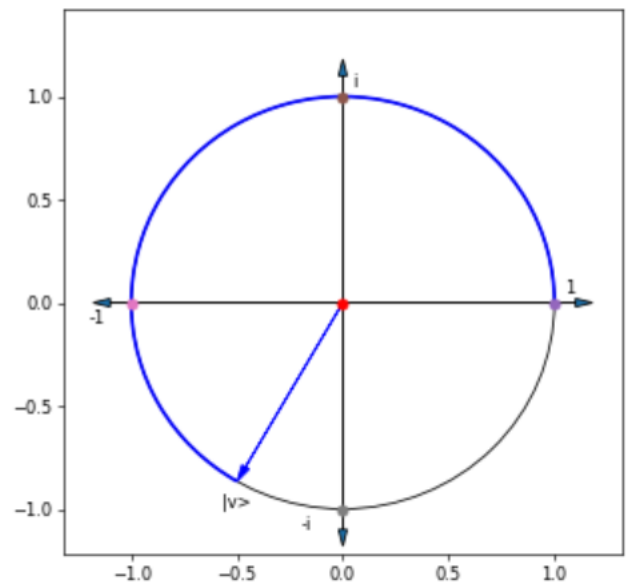
$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$$

Pure state on a Bloch sphere

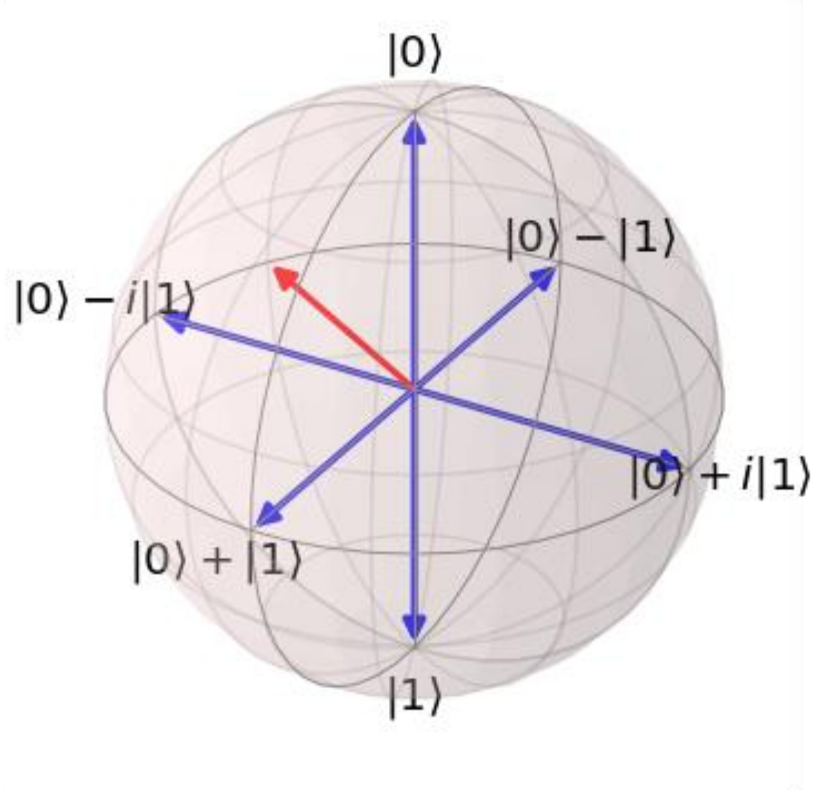
$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$$

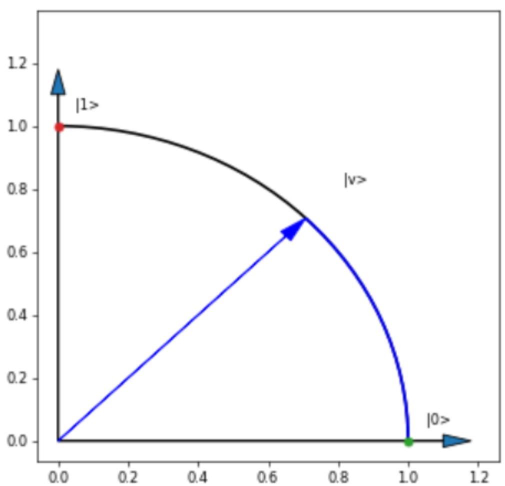
for $\theta = 90$ degrees, $\varphi = 240$ degrees

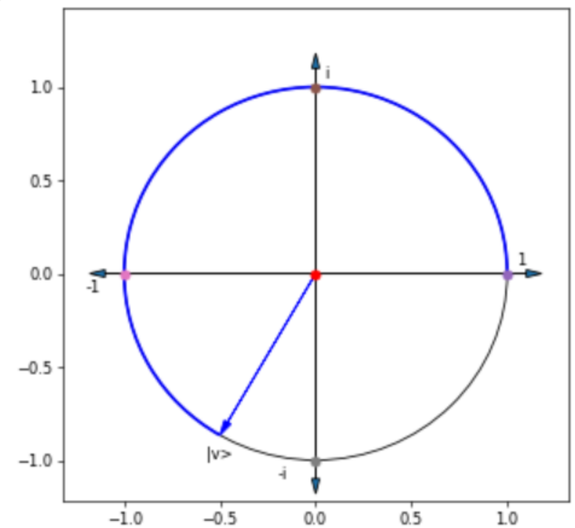




Pure state on a Bloch sphere







Density matrix

$|\psi\rangle = \mathrm{a}|0\rangle + \mathrm{b}|1\rangle$

$\rho = |\psi\rangle\langle\psi| = \binom{a}{b}(a^* \quad b^*) = \begin{pmatrix} aa^* & ab^* \\ ba^* & bb^* \end{pmatrix}$

Trace of density matrix is always equal to 1.

$$\rho^2 = |\psi\rangle\langle\psi||\psi\rangle\langle\psi| = \rho$$

If we have a mixed state, e.g., $|\psi\rangle = 0.3|\psi_1\rangle + 0.7|\psi_2\rangle$,then

$$\rho^2 \neq \rho$$

Trace of the mixed state can also be $\neq 1$.

Density matrix

$$|\psi\rangle = \cos\frac{\theta}{2}|0\rangle + e^{i\varphi}\sin\frac{\theta}{2}|1\rangle$$

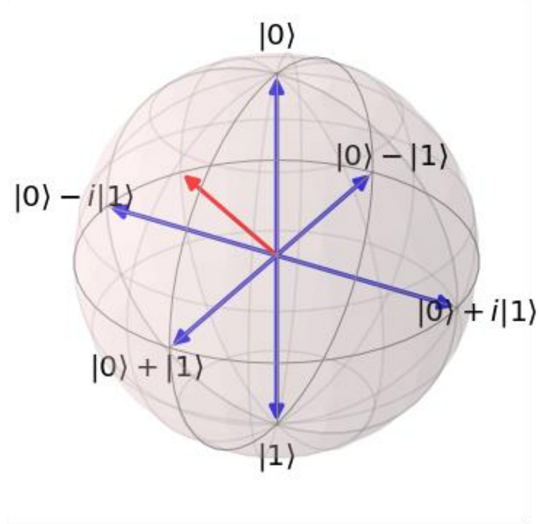
$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix} (\cos\frac{\theta}{2})^2 & \cos\frac{\theta}{2}\sin\frac{\theta}{2}e^{-i\varphi} \\ \cos\frac{\theta}{2}\sin\frac{\theta}{2}e^{i\varphi} & (\sin\frac{\theta}{2})^2 \end{pmatrix}$$

$$\rho = \frac{1}{2}(I + sin\theta cos\varphi \sigma_x + sin\theta sin\varphi \sigma_y + cos\varphi \sigma_z)$$

Density matrix

$$\rho = \frac{1}{2}(I + sin\theta cos\varphi\sigma_x + sin\theta sin\varphi\sigma_y + cos\varphi\sigma_z)$$

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$



Density matrix for classical prob

$|\psi\rangle = p_1|\psi_1\rangle + p_2|\psi_2\rangle$, where $|\psi_1\rangle = |0\rangle$ and $|\psi_2\rangle = |1\rangle$.

$$|0\rangle = \binom{1}{0}\begin{pmatrix}1 & 0\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$$

$$|1\rangle = \binom{0}{1}\begin{pmatrix}0 & 1\end{pmatrix} = \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}$$

Then our density matrix is:

$$\begin{pmatrix}p_1 & 0\\ 0 & p_2\end{pmatrix}$$

Density matrix on a Bloch sphere

For a given density matrix

$$\rho = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

alternative representation exists:

$$\rho = \frac{1}{2}(I + Tr(\rho\sigma_x)\sigma_x + Tr(\rho\sigma_y)\sigma_y + Tr(\rho\sigma_z)\sigma_z)$$

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

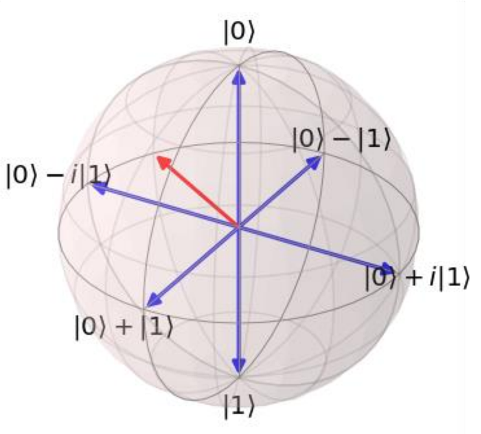
Density matrix on a Bloch sphere

$$\rho = \frac{1}{2}(I + Tr(\rho\sigma_x)\sigma_x + Tr(\rho\sigma_y)\sigma_y + Tr(\rho\sigma_z)\sigma_z)$$

Vector on a Bloch sphere will be with the following coordinates:

$$x = Tr(\rho\sigma_x), y = Tr\left(\rho\sigma_y\right), z = Tr(\rho\sigma_z)$$

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$



Prob state on a Bloch sphere

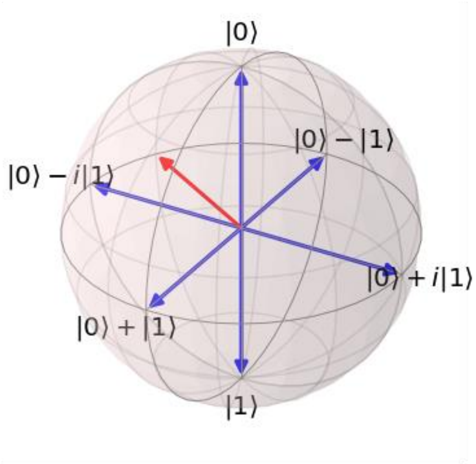
Given density matrix of a probabilistic state

$$\rho = \begin{pmatrix} p_1 & 0 \\ 0 & p_2 \end{pmatrix}$$

$$x = Tr(\rho\sigma_x) = 0, y = Tr(\rho\sigma_y) = 0,$$

$$z = Tr(\rho\sigma_z) = p_1 - p_2$$

On a Bloch sphere it will be a vector located on a Z-axis.



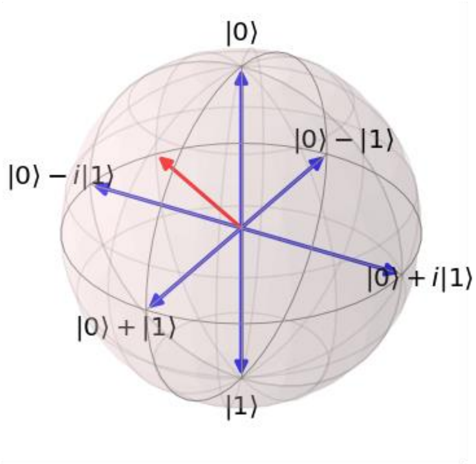
$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Mixed state on a Bloch sphere

$$\rho = \frac{1}{2}(I + Tr(\rho\sigma_x)\sigma_x + Tr(\rho\sigma_y)\sigma_y + Tr(\rho\sigma_z)\sigma_z)$$

$$x = Tr(\rho\sigma_x), y = Tr\left(\rho\sigma_y\right), z = Tr(\rho\sigma_z)$$

For a mixed state $\sqrt{x^2 + y^2 + z^2} < 1$, therefore, it will be located inside a Bloch sphere.



$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Fidelity

Fidelity measure

Fidelity is a measure of the "closeness" of two quantum states. It expresses the probability that one state will pass a test to identify as the other.

Given two density operators ρ and σ, the fidelity is generally defined as the quantity $F(\rho,\sigma) = (Tr(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}))^2$. In the special case where ρ and σ represent pure quantum states, the definition reduces to the squared overlap between the states: $F(\rho,\sigma) = |\langle\psi_\rho|\psi_\sigma\rangle|^2$

For pure states

$$F(\rho, \sigma) = \left|\langle\psi_\rho|\psi_\sigma\rangle\right|^2$$

Two vectors are orthogonal to each other if their inner product is zero. That means that the projection of one vector onto the other "collapses" to a point.

Therefore, two different states will have fidelity close to 0, but two similar states will have fidelity close to 1.

Fidelity vs fidelity distance

If we calculate fidelity, then fidelity distance can be obtained just by numeric operations:

$$d_F(\rho, \gamma) = \sqrt{2\left(1 - \sqrt{F(\rho, \gamma)}\right)}$$

Fidelity properties

- $0 \leq F(\rho, \sigma) \leq 1$
- $F(\rho, \sigma) = F(\sigma, \rho)$
- $F(\rho, \rho) = 1$
- $F(\rho, \sigma) = F\left(U\rho U^{\dagger}, U\sigma U^{\dagger}\right)$, for any unitary operator U. The fidelity is preserved by unitary evolution.

Fidelity properties

- If state ρ is pure, i.e., $\rho = |\psi\rangle\langle\psi|$, then $F(\rho,\sigma) = \langle\psi|\sigma|\psi\rangle$
- For one-qubit states $F(\rho,\sigma) = Tr(\rho\sigma) + 2\sqrt{\det(\rho)\det(\sigma)}$
- For pure one-qubit states ρ, $\det(\rho) = 0$, therefore, if at least one state is pure, then $F(\rho,\sigma) = Tr(\rho\sigma)$

Fidelity for pure-state qubits

$|\psi_1\rangle = \cos\frac{\theta_1}{2}|0\rangle + e^{i\varphi_1}\sin\frac{\theta_1}{2}|1\rangle$

$\rho = |\psi_1\rangle\langle\psi_1|$

$$= \begin{pmatrix} (\cos\frac{\theta_1}{2})^2 & \cos\frac{\theta_1}{2}\sin\frac{\theta_1}{2}e^{-i\varphi_1} \\ \cos\frac{\theta_1}{2}\sin\frac{\theta_1}{2}e^{i\varphi_1} & (\sin\frac{\theta_1}{2})^2 \end{pmatrix}$$

Similarly for $\sigma = |\psi_2\rangle\langle\psi_2|$, where $|\psi_2\rangle = \cos\frac{\theta_2}{2}|0\rangle + e^{i\varphi_2}\sin\frac{\theta_2}{2}|1\rangle$

Fidelity for pure-state qubits

$$\rho = \begin{pmatrix} (\cos\frac{\theta_1}{2})^2 & \cos\frac{\theta_1}{2}\sin\frac{\theta_1}{2}e^{-i\varphi_1} \\ \cos\frac{\theta_1}{2}\sin\frac{\theta_1}{2}e^{i\varphi_1} & (\sin\frac{\theta_1}{2})^2 \end{pmatrix}$$

$$F(\rho,\sigma) = Tr(\rho\sigma)$$

$$= (\cos\frac{\theta_1}{2})^2 \cdot \left(\cos\frac{\theta_2}{2}\right)^2 + \cos\frac{\theta_1}{2}\sin\frac{\theta_1}{2}e^{-i\varphi_1} \cdot \cos\frac{\theta_2}{2}\sin\frac{\theta_2}{2}e^{i\varphi_2}$$

$$+ \cos\frac{\theta_1}{2}\sin\frac{\theta_1}{2}e^{i\varphi_1} \cdot \cos\frac{\theta_2}{2}\sin\frac{\theta_2}{2}e^{-i\varphi_2} + (\sin\frac{\theta_1}{2})^2 \cdot (\sin\frac{\theta_2}{2})^2$$

Chapter 10 revision

Probabilities

$P(|\psi\rangle, x) = |c_x|^2$ - basic properties of ordinary probability.

$\langle\psi|\psi\rangle = \sum_x |c_x|^2 = 1$ (the sum of probabilities equals 1).

Rewriting probabilities

$$|c_x|^2 = |\langle\psi|x\rangle|^2 = \langle\psi|x\rangle\langle x|\psi\rangle$$

where $\Pi_x = |x\rangle\langle x|$ denotes the projection to the subspace spanned by $|x\rangle$.

$$x = \begin{array}{|c|c|} \multicolumn{1}{c}{m} & \multicolumn{1}{c}{n-m} \\ \hline y & z \\ \hline \end{array}$$

$$\mathbf{P}(|\psi\rangle, y) = \sum_z \mathbf{P}(|\psi\rangle, (y,z)) = \sum_z \langle\psi\,|y,z\rangle\langle y,z|\,\psi\rangle$$

$$= \langle\psi|\left(|y\rangle\langle y| \otimes I\right)|\psi\rangle = \langle\psi|\Pi_{\mathcal{M}}|\psi\rangle.$$

Orthogonal projection

Π_M denotes the operator of orthogonal projection onto the subspace $M = |y\rangle \otimes B^{\otimes(n-m)}$. Formula $P(|\psi\rangle, y)$=$\langle\psi|\Pi_M|\psi\rangle$ gives the definition of quantum probability also in the case where M is an arbitrary subspace. In this case the projection onto the subspace $M \subseteq N$ is given by the formula $\Pi_M = \sum_j |e_j\rangle\left\langle e_j\right|$, where e_j runs over an arbitrary orthonormal basis for M.

Probability comparisons

Classical probability	**Quantum probability**
Definition	
An event is a subset M of a fixed finite set N. A probability distribution is given by a function $w\colon N \to R$ with the properties a) $\sum_j w_j = 1$; b) $w_j \geq 0$. Probability: $\mathbf{Pr}(w, M) = \sum_{j \in M} w_j$.	An event is a subspace $\mathcal{M}$ of some finite-dimensional Hilbert space $\mathcal{N}$. A probability distribution is given by a state vector $\|\psi\rangle$, $\langle\psi\|\psi\rangle = 1$. Probability: $\mathbf{P}(\|\psi\rangle, \mathcal{M}) = \langle\psi\|\Pi_{\mathcal{M}}\|\psi\rangle$.
Properties	
1. If $M_1 \cap M_2 = \varnothing$, then $\mathbf{Pr}(w, M_1 \cup M_2) = \mathbf{Pr}(w, M_1) + \mathbf{Pr}(w, M_2)$.	1^{q}. If $\mathcal{M}_1 \perp \mathcal{M}_2$, then $\mathbf{P}(\|\psi\rangle, \mathcal{M}_1 \oplus \mathcal{M}_2) = \mathbf{P}(\|\psi\rangle, \mathcal{M}_1) + \mathbf{P}(\|\psi\rangle, \mathcal{M}_2)$.
2. (in the general case) $\mathbf{Pr}(w, M_1 \cup M_2) = \mathbf{Pr}(w, M_1) + \mathbf{Pr}(w, M_2) - \mathbf{Pr}(w, M_1 \cap M_2)$.	2^{q}. If $\Pi_{\mathcal{M}_1}\Pi_{\mathcal{M}_2} = \Pi_{\mathcal{M}_2}\Pi_{\mathcal{M}_1}$, then $\mathbf{P}(\|\psi\rangle, \mathcal{M}_1 + \mathcal{M}_2) = \mathbf{P}(\|\psi\rangle, \mathcal{M}_1) + \mathbf{P}(\|\psi\rangle, \mathcal{M}_2) - \mathbf{P}(\|\psi\rangle, \mathcal{M}_1 \cap \mathcal{M}_2)$.

Density matrix

$$\sum_k p_k \mathbf{P}(|\xi\rangle, \mathcal{M}) = \sum_k p_k \langle \xi_k | \Pi_{\mathcal{M}} | \xi_k \rangle$$

$$= \sum_k p_k \operatorname{Tr}\left(|\xi_k\rangle\langle\xi_k| \Pi_{\mathcal{M}} \right) = \operatorname{Tr}(\rho \Pi_{\mathcal{M}})$$

Here ρ denotes the density matrix $\rho = \sum_k p_k |\xi_k\rangle\langle\xi_k|$. The final expression here is what we take as the general definition of probability.

Density matrix

The operators of the form $\rho = \sum_k p_k \,|\xi_k\rangle\langle\xi_k|$ are precisely the Hermitian nonnegative operators with trace 1, i.e., operators that satisfy the conditions:

- $\rho = \rho^{\dagger}$
- $\forall|\eta\rangle \; \langle\eta|\rho|\eta\rangle \geq 0$
- $Tr\;\rho = 1$

From now on, by a density matrix we will mean an arbitrary operator with these properties.

Diagonal matrices

Diagonal matrices correspond to classical probability distributions on the set of basis vectors.

$$\rho = \sum_j w_j |j\rangle\langle j|$$

$$P(\rho, M) = \Pr(w, M)$$

$$\rho = \sum_j w_j \cdot (j)$$

Probability comparisons

Classical probability	Quantum probability
Properties	
3. Suppose a probability distribution of the form $w_{jk} = w_j^{(1)} w_k^{(2)}$ is specified on the set $N = N_1 \times N_2$. Consider two sets of outcomes, $M_1 \subseteq N_1$, $M_2 \subseteq N_2$. Then the probabilities multiply: $\mathbf{Pr}(w, M_1 \times M_2) = \mathbf{Pr}(w^{(1)}, M_1)\,\mathbf{Pr}(w^{(2)}, M_2)$.	3q. Suppose a density matrix of the form $\rho_1 \otimes \rho_2$ is defined on the space $\mathcal{N} = \mathcal{N}_1 \otimes \mathcal{N}_2$. Consider two subspaces, $\mathcal{M}_1 \subseteq \mathcal{N}_1$, $\mathcal{M}_2 \subseteq \mathcal{N}_2$. Then the probabilities likewise multiply: $\mathbf{P}(\rho_1 \otimes \rho_2, \mathcal{M}_1 \otimes \mathcal{M}_2) = \mathbf{P}(\rho_1, \mathcal{M}_1)\,\mathbf{P}(\rho_2, \mathcal{M}_2)$.
4. Consider a joint probability distribution on the set $N_1 \times N_2$. The event we are interested in does not depend on the outcome in the second set, i.e., $M = M_1 \times N_2$. The probability of such an event is expressed by a "projection" of the distribution onto the first set: $\mathbf{Pr}(w, M_1 \times N_2) = \mathbf{Pr}(w', M_1)$, where $w'_j = \sum_k w_{jk}$.	4q. In the quantum case, the restriction to one of the subsystems is described by taking a *partial trace* (see below). Thus, even if the initial state was pure, the resulting state of the subsystem may turn out to be mixed: $\mathbf{P}(\rho, \mathcal{M}_1 \otimes \mathcal{N}_2) = \mathbf{P}(\mathrm{Tr}_{\mathcal{N}_2} \rho, \mathcal{M}_1)$.

Partial trace

Let $X \in L(N_1 \otimes N_2) = L(N_1) \otimes L(N_2)$. The partial trace of the operator X over the space N_2 is defined as follows: if $X = \sum_m A_m \otimes B_m$, then $Tr_{N_2} X = \sum_m A_m (Tr B_m)$.

For 2-qubit systems, the partial trace is explicitly

$$\mathrm{Tr}_2 \begin{bmatrix} \rho_{00,00} & \rho_{00,01} & \rho_{00,10} & \rho_{00,11} \\ \rho_{01,00} & \rho_{01,01} & \rho_{01,10} & \rho_{01,11} \\ \rho_{10,00} & \rho_{10,01} & \rho_{10,10} & \rho_{10,11} \\ \rho_{11,00} & \rho_{11,01} & \rho_{11,10} & \rho_{11,11} \end{bmatrix} = \begin{bmatrix} \rho_{00,00} + \rho_{01,01} & \rho_{00,10} + \rho_{01,11} \\ \rho_{10,00} + \rho_{11,01} & \rho_{10,10} + \rho_{11,11} \end{bmatrix}$$

and

$$\mathrm{Tr}_1 \begin{bmatrix} \rho_{00,00} & \rho_{00,01} & \rho_{00,10} & \rho_{00,11} \\ \rho_{01,00} & \rho_{01,01} & \rho_{01,10} & \rho_{01,11} \\ \rho_{10,00} & \rho_{10,01} & \rho_{10,10} & \rho_{10,11} \\ \rho_{11,00} & \rho_{11,01} & \rho_{11,10} & \rho_{11,11} \end{bmatrix} = \begin{bmatrix} \rho_{00,00} + \rho_{10,10} & \rho_{00,01} + \rho_{10,11} \\ \rho_{01,00} + \rho_{11,10} & \rho_{01,01} + \rho_{11,11} \end{bmatrix}$$

Purification

An arbitrary mixed state $\rho \in L(N)$ can be represented as the partial trace $Tr_F|\psi\rangle\langle\psi|$ of a pure state of a larger system, $|\psi\rangle \in N \otimes F$. Such $|\psi\rangle$ is called a purification of ρ. (We may assume that $dimF = dimN$.)

Thank you for your attention!