Quantum Algorithms
Lecture 17
Quantum probability I

Zhejiang University

Probability for state vectors

Amplitudes

Let us discuss several “physical” aspects of quantum computation. Let a system of n qubits be in the state $|\psi\rangle = \sum_x c_x |x>$. The coefficients of the expansion relative to the classical basis are called amplitudes.

Measurement probability

The square of the modulus of the amplitude, $|c_x|^2$, equals the probability of finding the system in a given state x (compare with $P(|\psi\rangle, x)$=$|c_x|^2$). In other words, under a measurement of the state of this quantum system, a classical state will be obtained, according to the probability distribution $|c_x|^2$.

Probability

The quantity determined by formula $P(|\psi\rangle, x)=|c_x|^2$ possesses the basic properties of ordinary probability. The fact that the square of the modulus of the amplitude is the probability of observing the system in state x agrees with the fact that the physical states of quantum mechanics correspond to vectors of unit length, and transformations of these states do not change the length, i.e., they are unitary.

Sum of probabilities

Indeed, $\langle\psi|\psi\rangle = \sum_x |c_x|^2 = 1$ (the sum of probabilities equals 1), and the application of physically realizable operators must preserve this relation, i.e., the operator must be unitary.

Probability exceptions

Formula $P(|\psi\rangle, x) = |c_x|^2$ is sufficient for the definition of quantum computation and the class BQP. There are, however, situations for which this definition turns out to be inconvenient or inapplicable. Two fundamental examples are measurement operators and algorithms that are based on them, and the problem of constructing reliable quantum circuits from unreliable gates (error correction).

General definition

We, therefore, give a definition of quantum probability which generalizes both what we observe (the state of the system) and the result of the observation. We will arrive at this general definition by analyzing a series of examples.

Rewriting probability

To begin with, we rewrite the expression for the probability already obtained in the form

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

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

General definition

To make the next step toward the general definition of quantum probability, we compute the probability that the first m bits have a given value $y = (y_1, \ldots, y_m)$. Let us represent basis states in the form of two blocks:

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

We obtain:

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

Orthogonal projection

Here 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 $\Pi_M = \sum_j |e_j\rangle\langle e_j|$, where e_j runs over an arbitrary orthonormal basis for M.

Remark

The quantity $\sum_z |\langle F(x), z|U|x, 0^{N-n}\rangle|^2$, which appears in the definition of the evaluation of a function F: $B^n \to B^m$ by a quantum circuit, equals $P(U|x, 0^{N-n}\rangle, M)$, where $M = |F(x)\rangle \otimes B^{N-m}$. Recall once again the meaning of this definition: the circuit $U = U_L \cdots U_2 U_1$ computes F if for each x the probability to observe the correct result for $F(x)$ after application of the circuit to the initial state $|x, 0^{N-n}\rangle$ is at least $1 - \varepsilon$.

Projections

Projections do not represent physically realizable operators; more precisely, they do not describe the evolution of one state of a system to another over a fixed time period. Such evolution is described by unitary operators. Nonetheless, taking some liberty, it is possible to bestow physical meaning on projection. A projection selects a portion of system states from among all possible states.

Projections

Imagine a filter, i.e., a physical device which passes systems in states belonging to M but destroys the system if its state is orthogonal to M. (For example, a polarizer does this to photons.) If we submit a system in state $|\psi\rangle$ to the input of such a filter, then the system at the output will be in the state $|\xi\rangle = \Pi_M|\psi\rangle$. The probability associated to this state is generally smaller than one; it is $p = \langle\xi|\xi\rangle = \langle\psi|\Pi_M|\psi\rangle$. The number $1 - p$ determines the probability that the system will not pass through the filter.

Probability comparisons

Classical probability	**Quantum probability**
Definition	
An event is a subset M of a fixed finite set N.	An event is a subspace $\mathcal{M}$ of some finite-dimensional Hilbert space $\mathcal{N}$.
A probability distribution is given by a function $w\colon N \to R$ with the properties a) $\sum_j w_j = 1$; b) $w_j \geq 0$.	A probability distribution is given by a state vector $\lvert\psi\rangle$, $\langle\psi\rvert\psi\rangle = 1$.
Probability: $\mathbf{Pr}(w, M) = \sum_{j \in M} w_j$.	Probability: $\mathbf{P}(\lvert\psi\rangle, \mathcal{M}) = \langle\psi\rvert\Pi_{\mathcal{M}}\lvert\psi\rangle$.

Probability comparisons

Classical probability	Quantum probability
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 $\mathbf{P}(\lvert\psi\rangle, \mathcal{M}_1 \oplus \mathcal{M}_2) = \mathbf{P}(\lvert\psi\rangle, \mathcal{M}_1) + \mathbf{P}(\lvert\psi\rangle, \mathcal{M}_2)$.
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 $\mathbf{P}(\lvert\psi\rangle, \mathcal{M}_1 + \mathcal{M}_2) = \mathbf{P}(\lvert\psi\rangle, \mathcal{M}_1) + \mathbf{P}(\lvert\psi\rangle, \mathcal{M}_2) - \mathbf{P}(\lvert\psi\rangle, \mathcal{M}_1 \cap \mathcal{M}_2)$.

Note that the condition $M_1 \perp M_2$ (mutually exclusive events) is equivalent to the condition $\Pi_{M_1}\Pi_{M_2} = \Pi_{M_2}\Pi_{M_1} = 0$.

Non-additivity

If we have two nonorthogonal subspaces with zero intersection, the quantum probability is not necessarily additive. We give a simple example where $P(|\psi\rangle, M_1 + M_2) \neq P(|\psi\rangle, M_1) + P(|\psi\rangle, M_2)$.

Non-additivity

Let $|\xi\rangle = |0\rangle$, $M_1 = C(|0\rangle)$ (the linear subspace generated by the vector $|0\rangle$), $M_2 = C(|\eta\rangle)$, where $\langle\xi|\eta\rangle$ is close to 1. Then

$$1 = P(\quad |\xi\rangle, M_1 + M_2) \neq P(\quad |\xi\rangle, M_1) + P(\quad |\xi\rangle, M_2) \approx 1 + 1$$

Mixed states (density matrices)

Generalizing quantum object

Thus, we have defined, in the most general way, what quantity we measure. Now we need to generalize what object we perform the measurement on. Such objects will be something more general than state vectors or probability distributions. This will give us a definition of probability that generalizes both classical and quantum probability.

Probability to observe a state

Consider a probability distribution on a finite set of quantum states $\{|\xi_1\rangle, \ldots, |\xi_s\}$. The probability of the state $|\xi_j\rangle$ will be denoted by p_j; clearly $\sum_j p_j = 1$. We will calculate the probability of observing a state in the subspace M:

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

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

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.

Actually, this is an operator rather than a matrix, although the term “density matrix” is traditional. In the sequel, we will often have in mind a matrix, i.e., an operator expressed in a particular basis.

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.

Rank of a matrix

The maximum number of linearly independent rows in a matrix A is called the row rank of A, and the maximum number of linearly independent columns in A is called the column rank of A. For any matrix A, the row rank of A = the column rank of A.

Rank – examples

2×3 order matrix, $\mathbf{R} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \end{bmatrix}$

3 square submatrices:

$$\mathbf{R}_1 = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \quad \mathbf{R}_2 = \begin{bmatrix} 1 & 4 \\ 2 & 8 \end{bmatrix}, \quad \mathbf{R}_3 = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}$$

Each of these has a determinant of 0, so the rank is less than 2.
Thus the rank of $\mathbf{R}$ is 1.

Compute the ranks of the matrices

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 3 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & -1 \\ 2 & 2 & 0 \\ 1 & 3 & -2 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & -3 \\ 2 & 4 & -6 \\ 3 & 6 & -9 \end{pmatrix}$$

Answer.
3, 2, and 1.

For more examples: https://www.mathsisfun.com/algebra/matrix-rank.html

Pure and mixed states

The arguments about the “probability distribution on quantum states” were of an ancillary nature. The problem is how to generalize the notion of a quantum state to include classical probability distributions. The result we have obtained ($Tr(\rho\Pi_M)$) depends only on the density matrix, so that we may postulate that generalized quantum states and density matrices be the same. If a state is given by a density matrix of rank 1 (i.e., $\rho = |\xi\rangle\langle\xi|$), then it is said to be pure; if it is given by a general density matrix, it is called mixed.

Event probability

For a quantum state given by a density matrix ρ and a subspace M, the probability of the "event" M equals $P(\rho, M) = Tr(\rho \Pi_M)$.

Diagonal matrices

Diagonal matrices correspond to classical probability distributions on the set of basis vectors. Indeed, consider the quantum probability associated with the diagonal matrix $\rho = \sum_j w_j |j\rangle\langle j|$ and the subspace M spanned by a subset of basis vectors M. This probability can also be obtained by the classical formula: $P(\rho, M) = \Pr(w, M)$.

Diagonal matrices

From the physical point of view, a classical system is a quantum system that supports only diagonal density matrices (we will discuss the decoherence in the next section). A state of such a system may be denoted as

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

Mathematically, this is just a different notation of the probability distribution w. It is convenient when we need to simultaneously deal with classical and quantum systems.

Continuing comparison

Now we continue the comparison of the properties of classical and quantum probability; for the latter we shall now understand the general definition in terms of a density matrix.

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: $\mathbf{Pr}(w, M_1 \times M_2) = \mathbf{Pr}(w^{(1)}, M_1)\ \mathbf{Pr}(w^{(2)}, M_2)$.	3^{q}. 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: $\mathbf{P}(\rho_1 \otimes \rho_2, \mathcal{M}_1 \otimes \mathcal{M}_2) = \mathbf{P}(\rho_1, \mathcal{M}_1)\, \mathbf{P}(\rho_2, \mathcal{M}_2)$.

Probability comparisons

Classical probability	Quantum probability
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 $Tr_{N_2} X = \sum_m A_m \, (Tr B_m)$.

Choice of summands

Due to the universality property of the tensor product, the partial trace does not depend on the choice of summands in the representation $X = \sum_m A_m \otimes B_m$. This may seem somewhat obscure, so we will give a direct proof.

Proof

Let us choose orthonormal bases in the spaces N_1, N_2 and express the partial trace in terms of the matrix elements $X_{jj'kk'} = \langle j, j'|X|k, k'\rangle$. Let

$$A_m = \sum_{j,k} a^m_{jk} |j\rangle\langle k| \quad \text{and} \quad B_m = \sum_{j',k'} b^m_{j'k'} |j'\rangle\langle k'|$$

Proof

Then

$$X = \sum_{j,j',k,k'} X_{jj'kk'} |j,j'\rangle\langle k,k'| = \sum_m A_m \otimes B_m = \sum_{j,j',k,k',m} a^m_{jk} b^m_{j'k'} |j,j'\rangle\langle k,k'|$$

so the partial trace equals

$$\mathrm{Tr}_{\mathcal{N}_2} X = \sum_m \sum_{j,k} a^m_{jk} \left(\sum_l b^m_{ll} \right) |j\rangle\langle k| = \sum_{j,k} \sum_l X_{jlkl} |j\rangle\langle k|$$

Partial trace example

Examples: Partial trace (III): calculating matrices of partial traces

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

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

Partial trace example

Let us consider an example where taking the partial trace of the density matrix corresponding to a pure state leads to the density matrix corresponding to a mixed state.

Partial trace example

Let $N_1 = N_2 = B$ and $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$. In this case $\rho = \frac{1}{2}\sum_{a,b} |a,a\rangle\langle b,b|$, thus we obtain

$$Tr_{N_2}\rho = \frac{1}{2}\sum_{a} |a\rangle\langle a| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

Partial trace example

This matrix corresponds to a mixed state (pure states correspond to matrices of rank 1). Moreover, this mixed state is equivalent to a classical probability distribution: 0 and 1 have probabilities 1/2. Thus, discarding the second qubit yields a purely classical probability distribution on the first qubit.

$$Tr_{N_2}\rho = \frac{1}{2}\sum_a |a\rangle\langle a| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

Proposition – 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 $dimF = dimN$.)

Unitary purifications of quantum circuits

General quantum operations can be represented by unitary operations on larger systems.

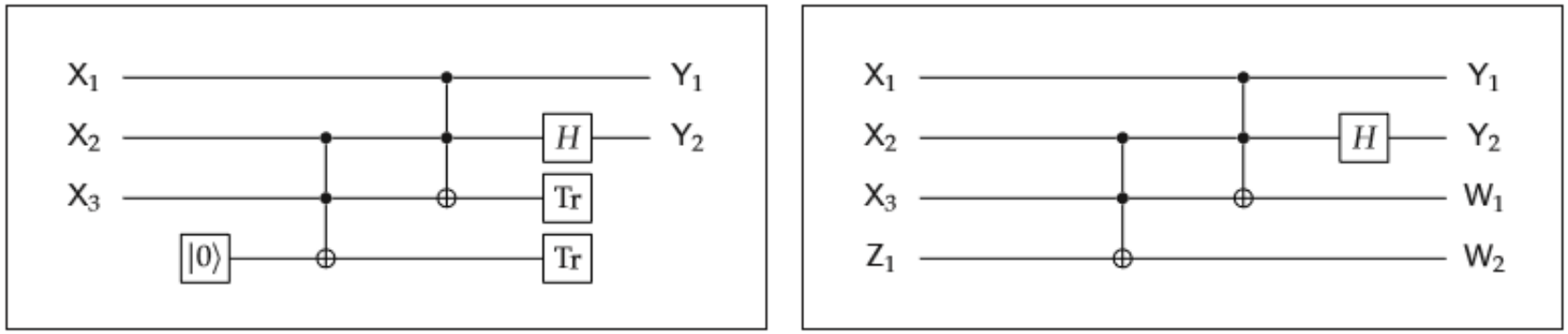

Figure 5: A general quantum circuit (left) and its unitary purification (right).

Purification – proof

Set $F = N^*$. Since ρ is a nonnegative (= positive semidefinite) Hermitian operator, there exists $\sqrt{\rho} \in L(N) = N \otimes N^*$. More explicitly, let us choose an orthonormal basis in which ρ is diagonal, i.e., $\rho = \sum_j p_j |\xi_j\rangle\langle\xi_j|$. Then $\sqrt{\rho} = \sum_j \sqrt{p_j} |\xi_j\rangle\langle\xi_j|$.

Purification – proof

Let us regard $\sqrt{\rho}$ as a vector of the space $N \otimes N^*$:

$$|\sqrt{\rho}\rangle = |\psi\rangle = \sum_j \sqrt{p_j}\,|\xi_j\rangle \otimes |\eta_j\rangle$$

where $|\eta_j\rangle = \langle\xi_j| \in N^*$.

This vector satisfies the desired requirements, i.e., $Tr_F(|\psi\rangle\langle\psi|) = \rho$. Indeed,

$$|\psi\rangle\langle\psi| = \sum_{jk} \sqrt{p_j p_k}\,(|\xi_j\rangle \otimes |\eta_j\rangle)(\langle\xi_k| \otimes \langle\eta_k|)$$

Purification – proof

Only terms with $j = k$ contribute to the partial trace. Therefore

$$Tr_{N^*}(|\psi\rangle\langle\psi|) = \sum_j p_j \left|\xi_j\right\rangle\left\langle\xi_j\right| = \rho$$

Schmidt decomposition

Consider a pure state $|\psi\rangle \in N \otimes F$. The so-called Schmidt decomposition holds:

$$|\psi\rangle = \sum_j \lambda_j |\xi_j\rangle \otimes |\eta_j\rangle$$

where $0 < \lambda_j \leq 1$, and the set of vectors $\{|\xi_j\rangle\} \subset N$ and $\{|\eta_j\rangle\} \subset F$ are orthonormal.

Schmidt decomposition

$$|\psi\rangle = \sum_j \lambda_j \, |\xi_j\rangle \otimes |\eta_j\rangle$$

Note that the numbers λ_j^2 are the nonzero eigenvalues of the partial traces $\rho = Tr_F(|\psi\rangle\langle\psi|)$ and $\rho' = Tr_N(|\psi\rangle\langle\psi|)$ (Hence the nonzero eigenvalues of ρ and ρ' coincide.) The number of such eigenvalues equals the rank of ρ and ρ'. For example, if $rank(\rho) = 1$, the Schmidt decomposition consists of one term, and vice versa.

Pure state

Thus the state $\rho = Tr_F(|\psi\rangle\langle\psi|)$ is pure if and only if $|\psi\rangle$ is a product state, i.e., $|\psi\rangle = |\xi\rangle \otimes |\eta\rangle$. In general, $rank(\rho)$ is the smallest dimension of the auxiliary space F which allows a purification of ρ.

Purification

Purification is unique up to unitary equivalence.

Let $|\psi_1\rangle, |\psi_2\rangle \in N \otimes F$ be two pure states such that $Tr_F(|\psi_1\rangle\langle\psi_1|) = Tr_F(|\psi_2\rangle\langle\psi_2|)$. Then $|\psi_2\rangle = (I_N \otimes U)|\psi_1\rangle$ for some unitary operator U on the space F.

Thank you for your attention!