Quantum spectral clustering based on quantum circuit model

Qingyu Li
(Dated: February 3, 2021)

In renct years, quantum machine learning is the most importance research field in quantum computing, which utilizes the quantum advantage to improve th performance of machine learning algorithms. As a vital task of machine leraning, clustering is uesd in many filed. Quantum clustering has also been taken seriously by researchers. However, lost of those quantum clustering algorithms only are theoretical and can't be easily described by quantum circuit model. Here, we present a quantum spectral clustering algorithm which bases on the quantum circuit model. We show that it only requires a logarithmic number of qubits and can achive a speed-up compared with the classical counterpart. And we perform numerical simulation to illustrate that the quantum spectral clustering algorithm has the same performance as the classical counterpart with logarithmic resourses and speed-up.

1. INTRODUCTION

Quantum machine learning is an emerging interdisciplinary research area which is the intersection of quantum mechanics and machine learning. In recent years, the research in quantum algorithms for machine learning problems has gained substantial momentum. Some of these algorithms include the application of quantum random walk to the community detection in quantum networks, quantum nearest neighbor methods for clustering problems, the deep learning in the context of quantum computing, and an accelerated unsupervised learning algorithm with the help of quantum based subroutines. Furthermore, quantum algorithms for topological and geometric analysis of data and quantum principal component analysis are introduced. The computational complexities of these algorithms are exponentially less than the classical algorithm when the data is accessible on a quantum RAM.

Clustering is one of the most widely used unsupervised machine learning algorithm for exploratory data analysis, with applications ranging from statistics, computer science, biology to social sciences or psychology. In virtually every scientific field dealing with empirical data, people attempt to get a first impression on their data by trying to identify groups of "similar behavior" in their data. The intuitive goal of clustering is to divide the data points into several groups such that points in the same group are similar and points in different groups are dissimilar to each other.

Compared to the "traditional algorithms" such as k-means or single linkage, spectral clustering has many fundamental advantages. With roots in graph theory, it uses the spectral properties of the Laplacian matrix to project the data in a low dimensional space where clustering is more efficient. Results obtained by spectral clustering often outperform the traditional approaches, spectral clustering is very simple to implement and can be solved efficiently by standard linear algebra methods.

In this paper, we propose a quantum spectral clustering algorithm. We use the phase estimation algorithm and amplitude amplification algorithm to quickly calculate the eigenvalues and eigenvectors of Laplace matrix, and then use the optimized measurement to cluster datapoints. Before our article, quantum spectral clustering has been studied once. But it has some problems. It doesn't use a mathematical model of spectral clustering but k clustering; it picks the wrong eigenvectors; it just proves that acceleration exists in special cases, not in general.

The remaining part of this paper is organized as follows: in the section II, there is a brief introduction to spectral clustering and K-means algorithm on the classical computers, including the algorithm itself and mathematical knowledge background. In the next section, the quantum spectral clustering will be introduced. Finally, the numerical simulations illustrate that the quantum spectral clustering can efficiently solve datapoints clustering problems.

2. THE CLASSICAL SPECTRAL CLUSTERING

A. The K-means Clustering

Spectral clustering algorithms are generally based on obtaining a clustering solution from the eigenvectors of a matrix which represents some form of a given data. It projects a higher-dimensional data vector into a lower-dimensional space where clustering is more efficient. In general, k-means algorithm is used to cluster the obtained low-dimensional data vectors. So we first give the description of this algorithm and then describe the spectral clustering.

Given a set of N data vectors, $v_1, v_2, \ldots, v_N$ k-means clustering tries to find best k centroids for assumed k number of clusters, $S_1, \ldots S_k$, by minimizing the following objective function:

$$\min(\sum_{c=1}^{k} \sum_{v_i \in S_c} ||v_i - m_c||^2) \tag{1}$$

Where m_c represent the center of the cluster S_C. And $||v_i - m_c||^2$ is the Euclidean distance measure between the data point v_i and the center m_c. The optimization

problem defined by the above obejective function is an NP-hard problem; nonetheless, it can be approximately minimized by using k-means algorithm, also known as Lloyd's algorithm.
The steps of this algoritm are as follows:
1) Initialize centroids for the clusters;
2) Assign each data point to the cluster with the closest centroid.
3)Assign the means of data in clusters as the new means: i.e., $m_c = \sum_{v_i \in S_c} \frac{v_i}{|S_c|}$
4)Repeat step 2 and 3 until there is no change in the means.

B. The Spectral Clustering

In this section, we try to briefly summarize the concept of the spectral clustering. Similarities between data points are most commonly represented by similarity graphs. i.e. undirected weighted graphs in which the vertices v_i and v_j are connected if the data points, x_i and x_j represented by these vertices are similar. And the weights on the edges w_{ij} indicates the amount of the similarity s_{ij} between x_i and x_j.

The construction of similarity graph $G(V, E)$ from a given data set $x_1, ..., x_N$ with pairwise similarities s_{ij} or distance d_{ij} can be done in many different ways. Three of the famous ones are: The undirected ϵ-neighborhood graph; The k-nearest neighborhood graph; The fully connected graph.

We usually use the similarity matrix to record the information of the similarity graph. By definition, the similarity matrix W is a adjacency matrix with the matrix elements, w_{ij} representing the weigh on the edge between vertex v_i and vertex v_j. The unnormalized Laplacian for the graph given by W is defined as:

$$L = W - D, \tag{2}$$

where D is the digonal degree matrix with diagonal elements $d_{ii} = \sum_{j=1}^{N} w_{ij}$. The smallest eigenvalues of the laplacian matrix is 0 and the elements of the associated eigenvector are equal to one. The multipliciy m of eigenvalue 0 gives the number of connected components in the graph. Clustering is generally done through the eigenvcetors associated to the first smallest eigenvalues $\lambda_1, ..., \lambda_k$ such that $\gamma_k = |\lambda_k - \lambda_{k+1}|$ gives the largest eigengap among all possible eigengaps.

Unormalized spectral clustering
Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.
1) Construct a similarity graph by one of the ways described in. Let W be its weighted adjacency matrix.
2) Compute the unnormalized Laplacian L.
3) **Compute the first k eigenvectors $u_1, \ldots, u_k$ of L.**
4) Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.
5) For $i = 1, \ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the $i-$th row of U.
6) Cluster the points $(y_i)_{i=1,\ldots,n}$ in $\mathbb{R}^k$ with the $k-$means algorithm into clusters $C_1, \ldots, C_k$.
Output: Clusters $A_1, \ldots, A_k$ with $A_i = \{j | y_j \in C_i\}$

3. QUANTUM SPECTRAL CLUSTERING BASED ON QUANTUM CIRCUIT MODEL

Suppose a given dataset M consists of $\{\boldsymbol{x}_{i=0,\ldots,N-1}\} \in \mathbb{R}^N$, generally where $N = 2^n$, $n \in \mathbb{Z}^+$. Clustering is the task of grouping objects in M in such a way that objects in the same cluster are more similar to each other than to those in ohter cluster. In general, there are varitious methods that measure the similar between two objects, such as euclidean distance. Clustering is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics and machine learning. Because of the importance of clustering and wideily applied, quantum clustering attach the interesting of scientist. Here, we introduce a quantum spectral clustering algorithm based on quantum circuit model. Our algorithm show the same performance as classical counterpart and has a speedup than classical counterpart. The quantum spectral clustring can be applied to many filed.

Now, we will show how to use quantum spectral clustering algorithm to cluster objects in M into k categories. The workflow of quantum spectral clustering algorithm is divided into three steps. First, we need to construct a similarity matrix or Laplacian matrix corresponding to the dataset M. Without loss of generality, we choose the mutual k-nearest neighbor graphy method to construct the similarity matrix W. If x_i and x_j are each other's k-nearest neighbor, so W_{ij} and W_{ji} both are 1. If not, W_{ij} and W_{ji} both are 0. Further, we can get the Laplacian matrix $L = D - W$, where D is a diagonal matrix $D_{ii} = \sum_j W_{ij}$. After, We get a k-sparse,positive definite and symmetric matrix L. Second, we will encode L into unitary operator $U = e^{-itL}$, and then we can calculate the eigenvalues and eigenstates of it. Last, we use quantum measruement to extract the result of clustering, the quantum measurement may repeat about N times. In the following article, we will step by step introduce the quantum spectral clustering algorithm in detail.

A. the quantum eigenproblem solver

Before we introduce how use quantum eigenproblem solver to get all eigenpairs of L. We first give a brief introduction of quantum pahse estimation algorithm. The quantum phase estimation algorithm which also referred to as quantum eigenvalue estimation algorithm is a quantum algorithm to estimate the phase of an given eigen-

vector of a unitary operator. More precisely, given a unitary operator $U = e^{-itH}$ and a eigenstates $|\psi\rangle$ such that $U|\psi\rangle = e^{2\pi i\theta}|\psi\rangle$, the algorithm estiamtes the value of θ with high probability within additive error ϵ, using $O(log(1/\epsilon))$ qubits and $O(1/\epsilon)$ controlled-U operations. In some article, people want to use the phase estimation algorithm to get the eigenvalues and eigenvectors of the interested unitray operator. The phase register is parperd $|0\rangle$ and the system register is initialized $\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}|j\rangle$ rather than a given eigenstate $|u_i\rangle$. The output of the variance of phase estimation algorithm is expected as $\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}\alpha_j|\widetilde{\lambda}_j\rangle|u_j\rangle$, where a_j is the overlap between eigenstate $|u_j\rangle$ and uniform superposition state $\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}|j\rangle$. However, if the overlap a_j is zero, the output does not contain the information of $|u_i\rangle$. For example, the eigenvectors of Pauli Z are $u_0 = |+\rangle$ and $u_1 = |-\rangle$. If the system register is uniform superposition $\frac{1}{\sqrt{2}}$, the overlap a_1 of u_1 and $\frac{1}{\sqrt{2}}$ are 0. To solve this problem, we introduce an enhanced quantum phase estiamtion algorithm. We increase the number of registers from 2 to 3, and prepare the second register (eigen register) and the third register (entangel register) to the maximum entangeled state $\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}|j\rangle|j\rangle$. We can proof that

$$\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}|j\rangle|j\rangle = \frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}|u_j\rangle|u_j^*\rangle \tag{3}$$

where $|u_i\rangle$ are eigenstates of U and $|u_j^*\rangle$ are conjugate of $|u_i\rangle$. And the $\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}|j\rangle|j\rangle$ can be prepared effectively with N Hadamard gate and N Control Not gate.

To get all eigenpairs of L, we need to encode the L into unitary operator $U = e^{-itL}$ as the phase estimation operator. The U only acts on the phase regieter and the eigen register. So, we can represent the enhanced phase estimation as

$$U_{pe}\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}|0\rangle|i\rangle|i\rangle = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}|\tilde{\lambda}_i\rangle|u_i\rangle|u_i^*\rangle. \tag{4}$$

The error $\epsilon = |\widetilde{\lambda}_i - \lambda_i|^2$ depends on the number of qubits in the phase register. Beacuse we don't care what the exact eigenvalues are, so if the error is small enough, we can ignore it.

Now we get the all eigenvalues and eigenvectors of L. The next step is to obtain the eigenvectors corresponding to the first k smallest eigenvalues. Suppose the first k eigenvalues $\lambda_{i=0,\ldots,k-1} \in Q$, where Q is a range $[0, s]$. In the amplitude amplification, we need an 'Black box' which can recognize whether a number is within a certain range. In particular, it can know the eigenvalues in Q. The Oracle can conveniently be represented by a function $f_O(x)$. By definition, $f_O(\lambda) = 1$ if $\lambda \in Q$ and $f_O(\lambda) = 0$ if $\lambda \notin Q$. By the Oracle, we can reform

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}|\tilde{\lambda}_i\rangle|u_j\rangle|u_j^*\rangle &= \frac{1}{\sqrt{N}}\sum_{\tilde{\lambda}_i\in Q}|\tilde{\lambda}_i\rangle|u_j\rangle|u_j^*\rangle + \frac{1}{\sqrt{N}}\sum_{\tilde{\lambda}_i\notin Q}|\tilde{\lambda}_i\rangle|u_j\rangle|u_j^*\rangle \\ &= \sqrt{\frac{k}{N}}|\alpha\rangle + \sqrt{\frac{N-k}{N}}|\beta\rangle.\end{aligned} \tag{5}$$

The iteration operator is

$$G = (2|\psi\rangle\langle\psi| - I)O, \tag{6}$$

where $|\psi\rangle = \frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}|\tilde{\lambda}_i\rangle|u_i\rangle|u_i^*\rangle$.

The effect of G can be understood in a beautiful way by realizing that the oracle operation O performs a reflection about the state $|\beta\rangle$ in the plan defined by $|\alpha\rangle$ and $|\beta\rangle$. That is, $O(a|\alpha\rangle + b|\beta\rangle) = (-a|\alpha\rangle + b|\beta\rangle)$. Similarly, $2|\psi\rangle\langle\psi| - I$ also performs a reflection in the plane defined by $|\alpha\rangle$ and $|\beta\rangle$, about the state $|\psi\rangle$. And the product of two reflcetion is a rotaion. The $R = O(\sqrt{N/k})$ iterations must be performed in order to obtain a superposition state of the firs k eigenvalues and corresponding eigenvectors with high probability. After enough iterations, the output state may not be exactly what we want. But the error can be ignored. By the iterations, suppose we get the

$$|\phi\rangle = \frac{1}{\sqrt{k}}\sum_{\tilde{\lambda}_i\in Q}|\tilde{\lambda}_i\rangle|u_i\rangle|u_i^*\rangle. \tag{7}$$

Since we don't need information about the eigenvalues, we need to untangle the state. So we apply the $U_{pe}^\dagger$ to the state $|\phi\rangle$. Then we obtain

$$U_{pe}^\dagger|\phi\rangle = |0\rangle\frac{1}{\sqrt{k}}\sum_{\tilde{\lambda}_i\in Q}|u_i\rangle|u_i^*\rangle. \tag{8}$$

The interesting part of the output state is $\frac{1}{\sqrt{k}}\sum_{\tilde{\lambda}_i\in Q}|u_i\rangle|u_i^*\rangle$. If we want get the clustering result, we need to measure the quantum state.

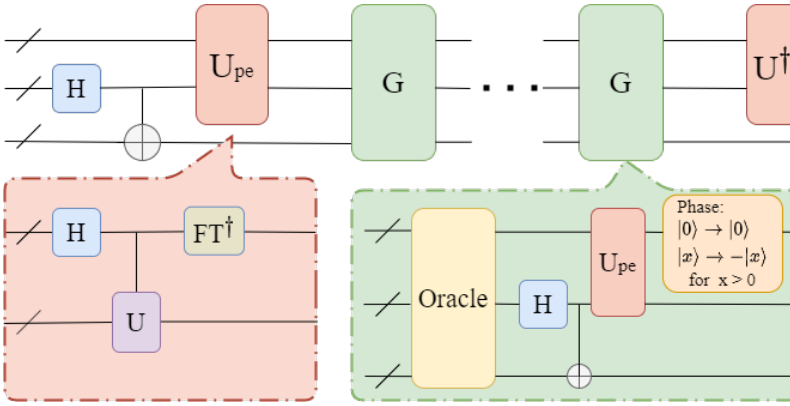


FIG. 1. The quantum circuit of quantum spectral clustering,$U = e^{2\pi iL}$, the *Oracle can recognize the first k eigenvalues*

B. Optimized measurement algorithm

Now, let's move on to second step. In classical K-means clustering sepctral relaxation is usually used to transform K-means clustering to the trace maximization problem. Given a set of M-dimensional data vectors $b_{i=1,..,N}$, the M-by-N data matrix B is composed of $[b_1, ..., b_N]$. Find the best partition of B can be reformed to find the optimal N-by-K indicator matrix X which satisfies

$$\max_{X^T X = I_k} trace(X^T B^T B X), \tag{9}$$

where K is the number of cluster. We can rewrite equation(9) as

$$\max_{X^T X = I_k} trace(B^T B X X^T), \tag{10}$$

where XX^T is a N-by-N hermitian matrix. In the process of spectral clustering, we need apply K-means to the column vector of $A = [u_1, ..., u_k]$. So let $B = A^T$, the second step is find the indicator matrix X which satisfies

$$\begin{aligned}\max_{X^T X = I_k} trace(\frac{1}{k} A A^T X X^T) &= \max_{X^T X = I_k} trace(\frac{1}{k} \sum_{i=1}^{k} |u_i\rangle\langle u_i| X \\ &= \max_{X^T X = I_k} trace(\rho_1 X X^T)\end{aligned} \tag{11}$$

If we can construct the observable $M = XX^T$ and the density operator $\rho = \frac{1}{k}\sum_{i=1}^{j} |u_i\rangle\langle u_i|$. The problem is transformed to find the maximum value of expectation of measurements. We can optimize the structure of observables M to achive the maximum value. The upper bound of max trace is the sum of the largest k eigenvalues of ρ_1. We can know, ρ_1's eigenvalues are $(n-k)$ 0 eigenvalues and k $\frac{1}{k}$ eigenvalues. So the maximum value of trace is $k\frac{1}{k} = 1$.

Classically, if we want to get the final clustering result, we need to fully know the eigenmatrix A or the covariance matrix AA^T, and then do k-means clustering for it. However, in quantum, if we want to know the eigenmatrix, we need to do quantum tomography on the quantum system. This is terrible. The cost of quantum tomography is expensive. So, what can we do to get the clustering information. We propose a feasible scheme to obtain the result of clustering by measurement. We have no prior knowledge of what kind of measurement operator will get the optimal result. Therefore, we need a set of methods to update the measurement operator and multiple measurements to ensure that we get an optimal solution. We use the local search algorithm to guide the measurement process. The objective function is

$$\max_{X^T X = I_k} trace(\rho_1 X X^T). \tag{12}$$

Our optimization strategy can be divided into the following steps. 1.Construct an initial indicator matrix X. and get $m = trace(\rho_1 X X^T)$.
2.Randomly select a row of X, swap the elements ,then get X's neighborhood X'. i.e.[0,1] is replaced [1,0].
3.if $m' = trace(\rho_1 X' X'^T) \geq m$, then update the $X = X'$. Repeat 2.
4.if $m' \leq m$. Repeat 2.

When we repaet the measurement enough times, the finally result X is approximation of the optimal solution $X*$. However, if we want to get the optimal solution, the cost is very high. In practice, we often do not take the optimal algorithm to get the optimal solution. Through the above algorithm, we can successfully extract the clustering information. it's complexity is lower than that of quantum tomography.

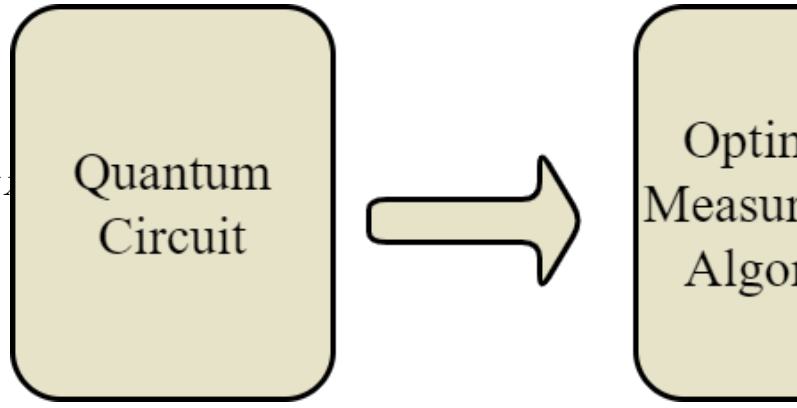
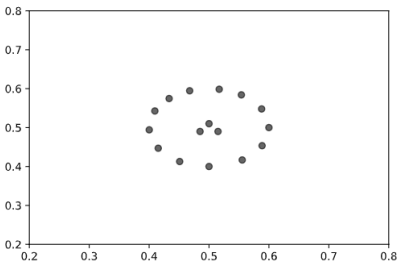


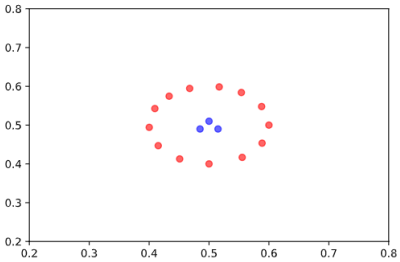
FIG. 2. The whole process of the quantum spectral clustering

4. NUMERICAL SIMULATION

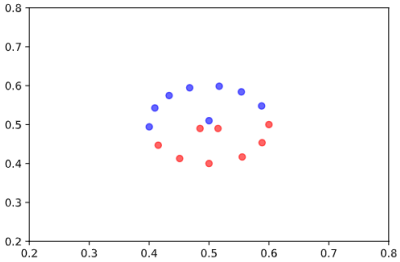
In order to verify that our algorithm is feasible, we present numerical simulations on simple synthetic datasets made of two concentric circles, as in the original work on spectral clustering, in order to benchmark the quality of the quantum algorithm. These simulation are made with a classical computer that simulates the quantum spectral clustering. Because of the performance limitations of classical computers, we cannot simulate higher-dimensional situations. In our simulation, the phase register has six qubits, the eigenstate register has four qubits, and the other four qubits are used to entangle with the eigenstate register. So the whole quantum system has 14 qubits and the number of points is $2^4 = 16$.



(a)initial graph

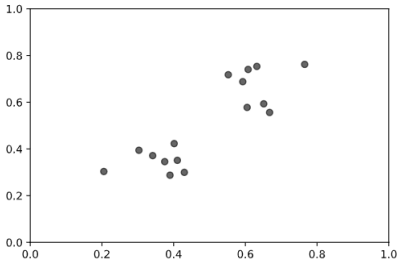


(b)quantum spectral cluster

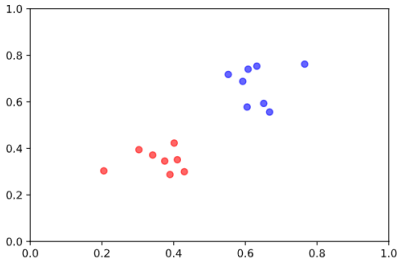


(c)k-means cluster

FIG. 3. quantum sepctral cluster



(a)initial graph



(b)quantum spectral cluster

FIG. 4. quantum sepctral cluster

5. APPENDIX

A. The maximum entangle state

Suppose we have a unitary operator U, if we want to get the eigenvalues and eigenvectors of U. We can use the phase estimation algorithm to do this. We prepare $|0\rangle\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}|i\rangle$ as the input state and construct the *control-U*. The output state is $\frac{1}{\sqrt{N}}\sum_{i=0}^{N-1}a_i|\lambda_i\rangle|u_i\rangle$, where λ_i and $|u_i\rangle$ are eigenvalues and eigenvectors of U respectively and a_i is the overlap between the equal superposition state and $|u_i\rangle$. However, there is a potential problem with this. If some $a_i = 0$, means the overlap between $|u_i\rangle$ and equal superposition state is zero, the output does not contain the information of $|u_i\rangle$.

We conclude that if we input the maximum entangled state, we will avoid this problem. The maximum entangle state $\sum_j|j\rangle|j\rangle$ consists of two register. We declare that the $\sum_j|j\rangle|j\rangle = \sum_j|u_j\rangle|u_j^*\rangle$, where $|u_j\rangle$ is the eigenvcetors of U. Suppose U is a normal matrix. And we can represent u_j as $\sum_{i=0}^{N-1}a_{i,j}|i\rangle$.

Now we give a detial proof of the argument above. First of all, let's prove a lemma. From the knowledge of matrix theory, we know the $\sum_j|u_j\rangle\langle u_j| = I$. And we suppose $|m\rangle$ as a arbitrary orthonormal basis. Thus, we can get

$$\begin{aligned}|m\rangle &= \sum_j|u_j\rangle\langle u_j||m\rangle \\ &= \sum_j\left[\sum_i a_{ij}|i\rangle\sum_k a_{kj}^*\langle k|\right]|m\rangle \\ &= \sum_j\sum_i\sum_k a_{ij}a_{kj}^*|i\rangle\langle k|m\rangle \\ &= \sum_j\sum_i a_{ij}a_{mj}^*|i\rangle.\end{aligned} \tag{13}$$

Then we can proof the final conclusion.

$$\sum_j|u_j\rangle|u_j^*\rangle = \sum_j\left[\sum_k a_{kj}|k\rangle\sum_i a_{ij}^*|i\rangle\right]$$