Quantum Persistent Homology for Time Series

Bernardo Ameneyro
Department of Mathematics
The University of Tennessee
Knoxville, TN 37996-1320, USA
bameneyr@vols.utk.edu

George Siopsis
Department of Physics and Astronomy
The University of Tennessee
Knoxville, TN 37996-1200, USA
gsiopsis@utk.edu

Vasileios Maroulas
Department of Mathematics
The University of Tennessee
Knoxville, TN 37996-1200, USA
vasileios.maroulas@utk.edu

Abstract—Persistent homology, a powerful mathematical tool for data analysis, summarizes the shape of data through tracking topological features across changes in different scales. Classical algorithms for persistent homology are often constrained by running times and memory requirements that grow exponentially on the number of data points. To surmount this problem, two quantum algorithms of persistent homology have been developed based on two different approaches. However, both of these quantum algorithms consider a data set in the form of a point cloud, which can be restrictive considering that many data sets come in the form of time series. In this paper, we alleviate this issue by establishing a quantum Takens’s delay embedding algorithm, which turns a time series into a point cloud by considering a pertinent embedding into a higher dimensional space. Having this quantum transformation of time series to point clouds, then one may use a quantum persistent homology algorithm to extract the topological features from the point cloud associated with the original time series.

Index Terms—quantum information, topological data analysis, machine learning, time series, Takens’s embedding theorem

1. INTRODUCTION

Topological data analysis (TDA) methods capture shape properties of data, and they have found a number of applications from biology [1]–[6] to chemistry and materials science [7]–[11], and from classification and clustering for action recognition [12], and handwriting analysis [13], to classification and clustering of signals [1], [12], [14]–[16]. Persistent homology is a powerful technique, on which TDA relies, and summarizes nonlinear and high-dimensional data retaining useful information about its shape.

The topological features discovered via persistent homology are displayed in a persistence diagram, and the efficient computation of such diagrams using packages such as Dionysus [17] and Ripser [18] leverage certain properties of simplicial complexes to create persistence diagrams efficiently [19]. However, these classical algorithms are not easily scalable and tend to deteriorate with the increase in the number of data size. Indeed, a point cloud consisting of \( n \) points possesses \( 2^n \) potential subsets that could contribute to the topology, so the best classical algorithm for estimating these features with accuracy \( \delta \) takes time \( O(2^n \log(1/\delta)) \) and requires \( O(2^n) \) bits to encode the subsets [20].

To bypass the computational bottleneck, quantum algorithms for homology (fixed data resolution/scale) [21], [22], and more recently, for persistent homology (varying data resolutions/scales) [23], [24] use a quantum random access memory (QRAM) to efficiently read a point cloud data set, and encode it into pertinent quantum states. One then uses these quantum states to build membership oracles that identify the features that are present at different scales. The main advantage of these algorithms is that they use relatively small QRAMs of size \( O(n^2) \), which only require \( O(\log n) \) calls to access the data [25], as opposed to classical algorithms that need memories of size \( O(2^{2n}) \).

These quantum algorithms make use of membership oracles to determine which objects are present at a certain scale. The oracles are simply subroutines that take into consideration a pertinent triangulation of point cloud data, called simplex and a positive number \( \epsilon \) encoded into qubits and return a qubit in state \( |1 \rangle \) if the simplex is present at scale \( \epsilon \), or a qubit in state \( |0 \rangle \) otherwise. On the other hand, the membership oracles that exist in the literature so far are merely restricted to simplicial complexes built from point cloud data sets.

In this paper, a quantum TDA algorithm is proposed by providing a membership oracle for time series. Indeed, a QRAM is queried to access the time series, and a quantum version of the Takens’s delay embedding theorem [26] is established for transforming a time series into a point cloud while preserving the relevant topological information. In turn, the oracle is plugged

This work has been partially supported by the NSF DMS-2012609, and DGE-2152168.
into the quantum algorithm for persistent homology [23] to study time series and their behavior (periodic, chaotic, etc.) by analyzing their corresponding point cloud.

This paper is organized as follows. Section II introduces the background information like the Takens’s delay embedding, and summarizes classical and quantum algorithms for TDA. Section III details our novel membership oracle as well as the quantum TDA algorithm for time series. Section IV shows two implementations of the quantum algorithm to time series data sets, and finally, Section V concludes with a brief discussion of our results.

II. BACKGROUND

A. Delay Embedding

Consider a time series \( x_t \) for \( t = 1, 2, ..., T \). The Takens’s delay embedding of the time series \( x_t \) is given by the vector

\[
v_i = (x_i, x_{i+\tau}, \ldots, x_{i+(d-1)\tau})
\]

for \( i = 1, ..., T - d \tau \), where \( d \) is the dimension of the point cloud, and \( \tau \) is the delay parameter. The Takens’s delay embedding theorem guarantees that the map from the time series to point cloud is an embedding and as such all the relevant topological information is retained [26]. The parametric choices of the dimension, \( d \), and the delay, \( \tau \), has been widely discussed in the literature, e.g. see [27] and references therein. For example, the time series is typically embedded into a \( d = 2, 3 \), space, and the delay parameter may be selected based on the autocorrelation.

B. Vietoris-Rips Filtration and Persistent Homology

An oriented \( k \)-simplex or simplex of dimension \( k \) is the collection of all convex combinations formed by \( (k+1) \) ordered and linearly independent vertices, a few examples are shown in Fig. 1 such as vertices, edges, triangles, tetrahedrons. They are written as the ordered list of their vertices, and a subset of this list is called a face of the simplex.

Homology studies simplicial complexes and chain complexes. A simplicial complex is a collection of simplices such that the faces of any simplex in the complex are also in the complex and the intersection of any two simplices in the complex is also a simplex in the complex. A commonly used simplicial complex is the Vietoris-Rips (VR) complex, which at scale \( \epsilon \) is defined by all the simplices with a diameter of at most \( \epsilon \). So, a simplex \( \sigma \) is in the complex \( S^\epsilon \) if and only if all the pairwise distances between its points are less than \( \epsilon \). On the other hand, the chain complex of \( S^\epsilon \) over \( \mathbb{Z}_3 = \{ -1, 0, 1 \} \) is the collection of formal sums of simplices in \( S^\epsilon \) with coefficients in \( \mathbb{Z}_3 \).

Classical algorithms encode the VR complex and chain complex, then compute its corresponding boundary matrix, and finally diagonalize it to extract topological features like the number of connected components, holes, voids, and \( k \) dimensional holes in general.

Algorithms for persistent homology consider instead a filtration or nested sequence of VR complexes \( S^\epsilon_0 \subseteq \cdots \subseteq S^\epsilon \) and analyze a boundary matrix in which the simplices are ordered according to the sequence [20]. Persistence diagrams are used to display the results of these algorithms. Points on the persistence diagram \((b, d)\) represent the number of \( k \) dimensional holes born at scale \( b \) that disappear at scale \( d \).

C. Quantum TDA Algorithms for point cloud data

The quantum algorithms appeared in [23], [24] established for the first time a quantum implementation for computing the persistent homology of a point cloud to reveal its topological properties across different scales. The study in [23] defines a linear map, the persistent Dirac operator, which tracks topological features across a pair of different scales. The persistent Dirac operator is a generalization of the Dirac operator established in the algorithms of [21], [22], which cannot track features across different scales/resolutions, and therefore they cannot obtain any persistent information. The work in [24] builds a block-encoding of the persistent combinatorial Laplacian and uses the quantum singular value transformation to extract the persistent topological features.

All of these quantum algorithms use a QRAM to store and access the data efficiently, they encode the points of a cloud as qubits and use superpositions to represent complexes. They also rely on a membership oracle to identify the simplices that are present at a certain scale and construct projections onto the corresponding VR complexes.
III. QUANTUM METHODOLOGY FOR TIME SERIES

A. Encoding Data and Simplices

Consider a discrete time series \( x_t, t = 1, 2, \ldots, T \), such that \( T > 0 \) is an integer. The data are stored in a QRAM, allowing for their quantum parallel access. The QRAM maps a quantum state \( |t⟩|0⟩ \) to the state \( |t⟩|x_t⟩ \) where \( |x_t⟩ \) is a state normalized to reflect the value \( x_t \). Such states can be encoded using \( O(\log_2 T) \) qubits.

Recall that the point cloud relates to the time series via the vectors \( v_i \) given in Eq. (1). We use \( T - \tau d \) qubits to encode these points so that any simplex \( \sigma \) formed with these points is represented by a quantum state \( |\sigma⟩ \), with the qubits corresponding to points in the simplex in state \( |1⟩ \), and all other qubits in state \( |0⟩ \). A balanced superposition of the states represents the maximal simplicial complex \( S \). Define \( S_d = \{-1, 0, 1\} \). Using these quantum states as a basis of the generated Hilbert space over \( S_d \) encodes the chain complex needed to extract the desired topological features.

B. Quantum Delay Embedding

Considering the time series data \( x_t \), along with the embedding parameters, the dimension, \( d \), of the space to which the times series is embedded, and the delay parameter, \( \tau \), one may check for membership in a simplicial complex by first computing the distances between points \( v_i \) in the embedded point cloud, defined in Eq. (1). The distance between two points \( v_i \) and \( v_j \) of the delay embedded time series is given by

\[
D(v_i, v_j) = \max_{0 \leq t < d} |x_{i+t\tau} - x_{j+t\tau}|. \tag{2}
\]

With the ability to create quantum states encoding the values \( x_t \), we can construct a quantum circuit that takes input states of the form \( |t_1⟩|t_2⟩|0⟩ \) and returns the output state \( |t_1⟩|t_2⟩|x_{t_1} - x_{t_2}⟩^2 \), where the last register contains an estimate of the difference between \( x_{t_1} \) and \( x_{t_2} \). To obtain an estimate with accuracy \( \delta \) it is necessary to make \( O(\delta^{-1}) \) calls to the QRAM and \( O(\delta^{-1}(\log_2 T)^2) \) [21], [25]. Moreover, this circuit can also operate in quantum parallel.

Notice that a simplex \( \sigma \) is in the VR complex \( S^\epsilon \) if and only if for each pair of vertices \( v_i \) and \( v_j \) in the simplex, \( D(v_i, v_j) \leq \epsilon \), where the distance, \( D \), is defined in Eq. (2). Being able to compute the difference \( |x_{t_1} - x_{t_2}| \) allows us to create an oracle \( O^\epsilon \) that verifies if the difference \( |x_{t_1} - x_{t_2}| \leq \epsilon \). Such an oracle takes as input the state \( |t_1⟩|t_2⟩|1⟩ \) and returns the state \( |t_1⟩|t_2⟩|\alpha^\epsilon⟩ \), where

\[
\alpha^\epsilon = \begin{cases} 
1 & \text{if } |x_{t_1} - x_{t_2}| \leq \epsilon \\
0 & \text{if } |x_{t_1} - x_{t_2}| > \epsilon.
\end{cases} \tag{3}
\]

Given a simplex \( \sigma \) encoded as the quantum state \( |\sigma⟩ \), where the qubits in state \( |1⟩ \) correspond to the vertices of \( \sigma \), we can make repeated calls to the oracle \( O^\epsilon \) to check if the simplex belongs to the VR complex at scale \( \epsilon \). In particular, we need to verify the inequality for all \( t_1 = i + \tau t \) and \( t_2 = j + \tau t \) with \( 0 \leq t < d \) and \( i < j \) such that \( v_i \) and \( v_j \) are vertices of \( \sigma \), where \( d \) and \( \tau \) as in Eq. (1) must be chosen according to the data. The vertices \( v_i, v_j \) are obtained from the quantum state \( |\sigma⟩ \), and they correspond to the qubits in state \( |1⟩ \). One concludes that \( \sigma \) is in \( S^\epsilon \) if all the calls to the oracle \( O^\epsilon \) yield an output with the last qubit in state \( |1⟩ \). However, if at least one of the calls returns an output with the last qubit in state \( |0⟩ \), it means the simplex is not yet present at that scale.

To that end, for fixed \( d \) and \( \tau \), a membership oracle \( O_{d, \tau}^\epsilon \) is constructed such that it acts on a quantum state \( |\sigma⟩|1⟩ \) according to

\[
O_{d, \tau}^\epsilon |\sigma⟩|1⟩ = \begin{cases} 
|\sigma⟩|1⟩ & \text{if } \sigma \in S^\epsilon \\
|\sigma⟩|0⟩ & \text{if } \sigma \notin S^\epsilon.
\end{cases} \tag{4}
\]

The membership oracle, \( O_{d, \tau}^\epsilon \), makes at most \( dk(k + 1)/2 \) calls to the oracle \( O^\epsilon \) in order to determine whether a simplex \( \sigma \) of dimension \( k \) is present in the VR complex at scale \( \epsilon \).

C. Quantum Algorithm for Persistent Homology

Having established a membership oracle \( O_{d, \tau}^\epsilon \), defined in Eq. (4), allows to consider a quantum persistent homology algorithm for time series. First, one uses \( T - \tau d \) qubits to represent the point cloud points, \( v_i \), defined in Eq. (1), which result from the embedding. Then, a simplex \( \sigma \) is represented by the quantum state \( |\sigma⟩ \), where the qubits corresponding to the vertices of \( \sigma \) are in state \( |1⟩ \), and all others are in state \( |0⟩ \). The boundary map \( \partial \) is essential to extract the topological features and it can be encoded using Pauli \( X \) operators, which are switches that change state \( |0⟩ \) into state \( |1⟩ \) and vice versa.

The operator given in Eq. (5) maps a simplex \( |\sigma_k⟩ \) of dimension \( k \) into a superposition of the \( (k-1) \)-simplices that conform its boundary, which are none other than the simplices obtained by removing each of the vertices in \( \sigma_k \).

\[
\partial_k |\sigma_k⟩ = \sum_{l=0}^{k} (-1)^l X_i |\sigma_k⟩. \tag{5}
\]
Given two fixed scales $\epsilon \leq \epsilon'$, the membership oracle in Eq. (4) is used to implement the projection operators, $P^\epsilon$ and $P'^\epsilon$, onto the subspaces that encode the VR complexes $S^\epsilon$ and $S'^\epsilon$. The boundary operator and the projections are in turn used to build a persistent Dirac operator

$$B^\epsilon,\epsilon' = \begin{pmatrix} -\xi I & \partial^\epsilon,\epsilon'_{k+1} & 0 \\ \partial^\epsilon,\epsilon'_{k} & \xi I & \partial^\epsilon,\epsilon'_{k+1} \\ 0 & \partial^\epsilon,\epsilon'_{k+1} & -\xi I \end{pmatrix},$$

(6)

where $\xi$ is an integer that is chosen according to the particular data at hand, and $\partial^\epsilon,\epsilon'_{k+1}$ is used to denote $P^\epsilon\partial_{k+1}P'^\epsilon$.

The parameter $\xi$ in Eq. (6) is an eigenvalue of the persistent Dirac operator and its eigenspace corresponds to the kernel of a generalization of the graph Laplacian [23]. In particular, the multiplicity of this eigenvalue is the persistent Betti number $\beta^\epsilon,\epsilon'$, which counts the number of $k$-dimensional holes in the point cloud that persist from scale $\epsilon$ up to at least scale $\epsilon'$.

To estimate a persistent Betti number, the quantum phase estimation algorithm is performed on the persistent Dirac operator in Eq. (6) and $N$ its dimension, while $l$ and $M$ are parameters to be chosen according to the data.

Next, one may use the Eq. (7) directly to estimate the persistent Betti numbers. Notice that if $l$ and $M$ are large enough, each of the terms in the sum in Eq. (7) is either equal to 1, if $p = l\lambda_s$, or very close to 0, if otherwise. Therefore $P(p)$ of Eq. (7) is proportional to the multiplicity of $\lambda_s$ at $p = l\lambda_s$ and 0 everywhere else. The desired persistent Betti number is then given by

$$\beta^\epsilon,\epsilon' = NP(|\xi|),$$

(8)

where the parameters $l$ and $M$ were chosen according to the eigenvalues of the persistent Dirac operators from Eq. (6). Indeed, the parameter $l$ was picked as the inverse of the gap between $\xi$ and the closest eigenvalue, then $M$ was chosen as a power of 2 larger than the greatest eigenvalue (in norm) of the persistent Dirac operator times $l$.

In order to obtain all persistent Betti numbers at various scales and represent them onto persistence diagrams, one needs to consider an increasing sequence of scales $\epsilon_0 < \ldots < \epsilon_n$, chosen according to the data. Then one implements the aforementioned quantum methodology to obtain the persistent Betti numbers for each pair $\epsilon_i, \epsilon_j$ with $0 \leq i \leq j \leq n$. Since the estimations are independent, this can be done in parallel. Finally, once we have all possible Betti numbers, and the scales, which appear and disappear, one may depict them onto persistence diagrams. Indeed, if $\mu^\epsilon,\epsilon'$ denotes the number of $k$-dimensional holes that appear in scale $\epsilon$ and disappear at scale $\epsilon'$, we have $\mu^\epsilon,\epsilon' = \beta^\epsilon,\epsilon' \beta^\epsilon,\epsilon' - (\beta^\epsilon,\epsilon' - \beta^\epsilon,\epsilon') - \beta^\epsilon,\epsilon' - \beta^\epsilon,\epsilon')$, where $\beta_k$ is defined in Eq. (8) for various scales.

IV. RESULTS

Two time series are considered to discover their shape properties. The first data is a simulated periodic discrete time series, and the second is a segment from an electroencephalography (EEG) measurement of a brain taken while the subject listened to music [28]. In both cases the task is to employ the quantum framework of Section III by first embedding the time series into a point cloud using the quantum delay embedding, and then examining the topological features of the corresponding embedded point clouds using the quantum persistent homology algorithm. Precisely, we simulate the quantum states and membership oracle Eq. (4) to construct a matrix representation of the persistent Dirac operator Eq. (6). Then we obtain the eigenvalues of this matrix and use Eq. (7) to estimate the persistent Betti numbers.

A. Periodic time series example

Consider the discrete time series based on the periodic function $\sin(2\pi t)$. Using a delay $d = 2$, $\tau = 1$, the time series is embedded into the 2-dimensional point cloud as shown in Fig. 2. Notice that this function has one

![Fig. 2](image-url)
Fig. 3: Vietoris Rips complexes of the embedded point cloud associated with the sinusoidal signal of Fig. 2 at scales (a) $\epsilon_1 = 1.2$, the points are pairwise connected, which results in a hole in the middle; and (b) $\epsilon_2 = 2.2$, the points become totally connected and the hole disappears. The blue squares are the balls of diameter $\epsilon_i$ around each point.

Fig. 4: Persistence diagram for the time series in Fig 2. The horizontal axis marks the scales at which the topological features are born, while the vertical axis marks the scales at which they disappear. The size of the dots represents the number of features that appear and disappear at the same scales. Finally, the orange dots along the vertical axis represent the connected components, while the blue dots closer to the diagonal line are the one dimensional holes.

maximum (or minimum) per period, which translates to one hole in the corresponding embedded point cloud, constructed by the quantum Taken’s delay embedding.

To construct the persistence diagram, an increasing sequence of twenty-five different scales starting at $\epsilon_0 = 0.0 < \epsilon_1 = 0.1 < \cdots < \epsilon_{24} = 2.4$ with step size 0.1 taken into account. Using the quantum persistent homology algorithm, the persistent Betti numbers of dimensions 0 and 1 for each possible pair of scales are computed, and depicted in Fig. 4. As expected, a single one dimensional hole exists due to periodicity, and this is born at scale 1 and died at scale 2. On the other hand, the number of connected components, which is initially four, is reduced to one when the one dimensional hole is formed at scale 1.

B. An electroencephalogram example

A segment of fifty measurements from an electroencephalogram (EEG) signal while the subject listened to music is considered in Fig. 5.

The point cloud obtained using the quantum delay embedding algorithm for a delay $\tau = 8$ and $d = 2$ is shown in Fig. 5. Next, the persistence diagrams are constructed using the quantum persistent homology algorithm for an increasing sequence of scales $\epsilon_k = k$, for $k = 0, \ldots, 15$. The persistence diagram of dimensions 0 (connected components) and 1 (holes) are depicted in Fig. 6. In this case one may see a number of small one dimensional holes that appear and disappear very quickly indicating that the segment of the signal is not periodic. The number of connected components is reduced quickly as shown by the large circles near the origin. Nevertheless, a few survive past scale 6 which indicates the presence of clusters in Fig. 5 (b).

V. CONCLUSION AND DISCUSSION

This paper makes possible for the first time, the topological data analysis of a time series by providing a quantum delay embedding of a time series into a point cloud. We introduced a new membership oracle for time series data, by changing the way in which the data is stored in the QRAM, and how it is used in the quantum topological data analysis algorithms to construct the relevant linear operators. The membership oracle provided here is for Vietoris-Rips complexes, but it is possible that a similar approach could be extended for other types...
of complexes dependent on the data. For example, lazy witness complexes are similar to VR complexes and retain the same topological characteristics, but in order to check for membership it suffices to consider only a subset of the vertices. This could further reduce the amount of resources that the quantum algorithm requires to extract the persistence information, and we plan to examine this in the future. Additionally, we would like to do a full scale complexity analysis for the quantum framework for persistent homology of time series.

**References**

[1] S. Emrani, T. Gentimis, and H. Krim, “Persistent homology of delay embeddings and its application to wheeze detection,” *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 459–463, 2014.

[2] C. M. Pereira and R. F. de Mello, “Persistent homology for time series and spatial data clustering,” *Expert Systems with Applications*, vol. 42, no. 15, pp. 6026–6038, 2015.

[3] M. Nicolau, A. J. Levine, and G. Carlsson, “Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 17, pp. 7265–7270, 2011.

[4] J. Mike and V. Maroulas, “Combinatorial hodge theory for equitable kidney paired donation,” *Foundations of Data Science*, vol. 1, no. 1, pp. 87–101, 2019.

[5] J. Mike, C. D. Sumrall, V. Maroulas, and F. Schwartz, “Nonlandmark classification in paleontology: computational geometry as a tool for species discrimination,” *Paleobiology*, pp. 1–11, 2016.

[6] V. Maroulas, C. P. Micucci, and F. Nasrin, “Bayesian topological learning for classifying the structure of biological networks,” *Bayesian Analysis*, vol. 17, no. 3, pp. 711 – 736, 2022.