Descriptors of Energy Landscapes using Topological Analysis

Henry Adams, Colorado State University
DELTA NSF #1934725
Descriptors of Energy Landscapes using Topological Analysis

- An introduction to persistent homology
- The n-alkane energy landscapes
- Machine learning tasks
Descriptors of Energy Landscapes Using Topological Analysis

3N Energy Landscape (Simulation/Experiment) → Dimensionality Reduction → Topology of Reduced Energy Landscapes → Predictive Machine Learning → Accelerated Sampling → Optimized Synthetic Conditions, Phase Behavior, Tuning Catalytic Pathways

PCA, Non-linear Methods, Generalized Collective Coordinates

Morse Theory, Persistent Homology, Catastrophe Theory, Singularity Theory

Energy

NSF #1934725
Abstract: Many of the properties of a chemical system are described by its energy landscape, a real-valued function defined on a high-dimensional domain. I will explain how topology, and in particular persistent homology, can be used in order to describe some of the pertinent features of an energy landscape. As a motivating example, the low-energy conformation space of an aluminate molecular anion interacting with a simple potassium cation can be well-modeled by the permutohedron (joint work with Aurora Clark and Hung Le). In a recently funded NSF Harnessing the Data Revolution project, the DELTA team is learning how to identify and leverage changing topological features of energy landscapes across a range of chemical conditions in order to predict reactivity.
Morse Theory Overview
Morse Theory Overview

Index 0 critical point  
Index 1 critical point  
Index 2 critical point

Images from Wikipedia
Fig. 16. Interaction energy between glucose and ethane under the three translational degrees of freedom. Left: isosurface of the electrostatic interaction pseudo-colored with the corresponding van der Waals potential. Middle: full MS complex with 564 critical points. Right: simplified MS complex with 166 critical points highlighting good candidate binding sites.

*Topological hierarchy for functions on triangulated surfaces* by Peer Timo-Brener, Herbert Edelsbrunner, Bernd Hamann, and Valerio Pasucci
An Introduction to Persistent Homology
$i$-dimensional homology $H_i$ “counts the number of $i$-dimensional holes”

- Six connected components
  - No 1-dimensional holes
  - No 2-dimensional holes

- One connected component
  - No 1-dimensional holes
  - One 2-dimensional hole

- One connected component
  - Three 1-dimensional holes
  - No 2-dimensional holes

- One connected component
  - Two 1-dimensional holes
  - One 2-dimensional hole
Input: Real-valued function on a space. Output: barcode.
Input: Real-valued function on a space. Output: barcode.
Sublevelset persistent homology

EL (PageRank) of ion-pair formation of Na$^+$ and OH$: sublevel sets

Input: Real-valued function on a space. Output: barcode.
Sublevelset persistent homology

EL (PageRank) of ion-pair formation of Na\(^+\) and OH\(^-\): sublevel sets

Input: Real-valued function on a space. Output: barcode.
Sublevelset persistent homology

EL (PageRank) of ion-pair formation of $\text{Na}^+$ and $\text{OH}^-$: sublevel sets

Input: Real-valued function on a space. Output: barcode.
Input: Real-valued function on a space. Output: barcode.
Sublevelset persistent homology

EL (PageRank) of ion-pair formation of Na\(^+\) and OH\(^-\): sublevel sets

Input: Real-valued function on a space. Output: barcode.
4.1 Persistent Homology

Persistent homology is a technique for analyzing the topology of a space over a range of scales. It is based on the concept of filtration parameters, which are used to construct a sequence of nested simplicial complexes. This is achieved by defining a filtration parameter that increases with time. As the parameter grows, simplices are added to the complex, reflecting the growth of topological features in the underlying space.

For example, consider a sequence of nested Vietoris–Rips complexes, with scale parameter \( r \), and a fixed dataset \( X \). The sequence \( \{VR(r)\} \) denotes a metric space, and the choice of scale parameter \( r \) determines the size of the neighborhoods used to construct the complex.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]

The horizontal axis in Figure 4 contains the varying \( t \)-values and display the result as a persistent homology barcode. See Figure 4. The idea of persistent homology is that long intervals in the persistence barcodes correspond to real topological features of the underlying space. We disregard short intervals as noise. Hence, this barcode reflects the fact that our points intersect the vertical line through scale \( r \) map.

The topological profile of this example, shown in Figure 3, contains four nested Vietoris–Rips complexes, with \( r \) increasing from left to right. The black dots denote a metric space, and the inclusion \( VR(r) \) is a functor. This means that for \( t \), which is the rank of the \( k \)-th homology group, we have

\[
H_k(X, t) \cong H_k(VR(r(t)), r(t)) \quad \text{for all } t.
\]
Persistent homology


Significant features persist.
Persistent homology applied to data
Example: Cyclo-Octane (C₈H₁₆) data
6040 points in 24-dimensional space

Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.
Persistent homology applied to data
Example: Cyclo-Octane (C$_8$H$_{16}$) data
6040 points in 24-dimensional space

Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.
Persistent homology applied to data
Example: Cyclo-Octane ($\text{C}_8\text{H}_{16}$) data
6040 points in 24-dimensional space

Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.
Sublevelset persistent homology of the $n$-alkanes

Joint with Johnathan Bush, Mark Heim, Joshua Mirth, Yang Zang, Yanqin Zhai, and the DELTA team
Sublevelset persistent homology of the $n$-alkanes
Sublevelset persistent homology of the $n$-alkanes

We have a complete characterization of the persistence barcodes of all $n$-alkanes.
Sublevelset persistent homology of the $n$-alkanes

We have a complete characterization of the persistence barcodes of all $n$-alkanes.
Sublevelset persistent homology of the $n$-alkanes

We have a complete characterization of the persistence barcodes of all $n$-alkanes
We have a complete characterization of the persistence barcodes of all \( n \)-alkanes.
Sublevelset persistent homology of the $n$-alkanes

We have a complete characterization of the persistence barcodes of all $n$-alkanes
Sublevelset persistent homology of the $n$-alkanes

We have a complete characterization of the persistence barcodes of all $n$-alkanes.
Sublevelset persistent homology of the $n$-alkanes

We have a complete characterization of the persistence barcodes of all $n$-alkanes.
We have a complete characterization of the persistence barcodes of all $n$-alkanes.
We have a complete characterization of the persistence barcodes of all $n$-alkanes.
Sublevelset persistent homology of the $n$-alkanes

We have a complete characterization of the persistence barcodes of all $n$-alkanes
We have a complete characterization of the persistence barcodes of all $n$-alkanes.
We have a complete characterization of the persistence barcodes of all $n$-alkanes.
We have a complete characterization of the persistence barcodes of all $n$-alkanes.
Sublevelset persistent homology of heptane

We have a complete characterization of the persistence barcodes of all $n$-alkanes.
Sublevelset persistent homology of octane

We have a complete characterization of the persistence barcodes of all $n$-alkanes
We have a complete characterization of the persistence barcodes of all $n$-alkanes.
Sublevel set persistent homology of the \( n \)-alkanes

- We have a formula for the birth and death time of each bar in every homological dimension.

- The proof uses a Künneth formula for persistent homology by Jose Perea and Hitesh Gakhar.

We have a complete characterization of the persistence barcodes of all \( n \)-alkanes.
We have a complete characterization of the persistence barcodes of all $n$-alkanes.
A notion of distance between energy landscapes

Bottleneck and Wasserstein distances are “edit” distances between barcodes
A notion of distance between energy landscapes

Bottleneck and Wasserstein distances are “edit” distances between barcodes
The cyclo-alkanes

- We have a plan in place for computing the sublevelset persistent homology of cyclo-octane, C₈H₁₆, using Vietoris-Rips complexes and the lower star filtration.
The cyclo-alkanes

- We have a plan in place for computing the sublevelset persistent homology of cyclo-octane, C\textsubscript{8}H\textsubscript{16}, using Vietoris-Rips complexes and the lower star filtration.

Non-Manifold Surface Reconstruction from High Dimensional Point Cloud Data
by Shawn Martin and Jean-Paul Watson, 2010.
Machine learning tasks

• Big goal: Predict how the energy landscape of a chemical system evolves as a result to changes in salinity, temperature, pressure, etc.

• Related subgoal: Predict how the sublevelset persistent homology evolves as a result to changes in salinity, temperature pressure, etc.
  • Halide task

\[
\text{H} - \text{C} - \text{C} - \text{Br} \\
\text{H} \quad \text{H}
\]

• Big goal: Given a partial sampling of a restricted energy landscape, predict the energy value at a new sample point.

• Related subgoal: Given a partial sampling of a restricted energy landscape, predict the sublevelset persistent homology.