Topology of Discrete Sampled Signals

Daniel Coble

July 4, 2023

Contents

1 Introduction

This document discusses topological methods of analyzing signals. Through multiple derivations we connect topological methods with more classical methods of signal processing. The goals of this document are to:

1. Investigate the 'meaning' of topological features and discover topological schemes wherein features carry the most information.

2. Improve the computational efficiency of topological feature extraction.

1.1 Preliminaries

Let a signal $x(t)$ be defined as a period signal defined between 0 and 2π. The discrete sampling of this signal will be evenly is $x(2\pi k/N)$ $(0 \leq n < N)$ and will be denoted $x(k)$ or simply the vector x. We also denote a sample of x as x_i and a subsequence from i to j as $x_{i:j}$. The first step of TDA is to develop the embedding

$$f : \mathbb{R} \to \mathbb{R}^d \tag{1}$$

$$x(t) \to x(t) \times x(t-\tau) \times x(t-2\tau) \times ... \times x(t-(d-1)\tau) \tag{2}$$

This embedding creates a continuous topology, and the embedding of the sampled $x(k)$ produces a discrete topology in the embedding.

1.2 Discussion of Results

2 Pure Sinusoid

Say $x(k) = A\sin(2\pi k/N)$. Consider topological dimension d. The proof proceeds by repeated use of trigonometric rules. At (12), a rule for summation of equally sampled cosines is applied, but the proof is omitted. With ε being the distance between two embedded points, we have:

$$\begin{aligned}
\varepsilon^2 =& \|x_{i:i+d} - x_{j:j+d}\|_2^2 && (3)\\
=& \sum_{k=0}^{d-1} (x_{i+k} - x_{j+k})^2 && (4)\\
=& \sum_{k=0}^{d-1} \left(A\sin\left(\frac{2\pi(i+k)}{N}\right) - A\sin\left(\frac{2\pi(j+k)}{N}\right)\right)^2 && (5)\\
=& A^2 \sum_{k=0}^{d-1} \left(\sin\left(\frac{2\pi(i+k)}{N}\right) - \sin\left(\frac{2\pi(j+k)}{N}\right)\right)^2 && (6)\\
=& A^2 \sum_{k=0}^{d-1} \left(2\cos\left(\frac{2\pi(i+k)}{N} + \frac{2\pi(j+k)}{N}\right)\sin\left(\frac{2\pi(i+k)}{N} - \frac{2\pi(j+k)}{N}\right)\right)^2 && (7)\\
=& 4A^2 \sum_{k=0}^{d-1} \cos^2\left(\frac{2\pi(i+j+2k)}{N}\right)\sin^2\left(\frac{2\pi(i-j)}{N}\right) && (8)\\
=& 4A^2 \sin^2\left(\frac{2\pi(i-j)}{N}\right) \sum_{k=0}^{d-1} \cos^2\left(\frac{2\pi(i+j+2k)}{N}\right) && (9)\\
=& 4A^2 \sin^2\left(\frac{2\pi(i-j)}{N}\right) \sum_{k=0}^{d-1} \frac{1}{2}\left(1+\cos\left(\frac{4\pi(i+j+2k)}{N}\right)\right) && (10)\\
=& 2A^2 \sin^2\left(\frac{2\pi(i-j)}{N}\right) \sum_{k=0}^{d-1} \left(1+\cos\left(\frac{4\pi(i+j+2k)}{N}\right)\right) && (11)\\
=& 2A^2 \sin^2\left(\frac{2\pi(i-j)}{N}\right) \left(d+\cos\left(\frac{4\pi(i+j+d-1)}{N}\right)\frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right) && (12)
\end{aligned}$$

The H_1 group is born when all adjacent points are within ε_1 of each other. So we examine the ε^2 function for $i-j$=1 and find the maximum

$$\begin{aligned}
\varepsilon_1^2 =& \max_{i-j=1} 2A^2 \sin^2\left(\frac{2\pi(i-j)}{N}\right)\left(d+\cos\left(\frac{4\pi(i+j+d-1)}{N}\right)\frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right) && (13)\\
=& 2A^2 \max_i \quad \sin^2\left(\frac{2\pi}{N}\right)\left(d+\cos\left(\frac{4\pi(2i+d)}{N}\right)\frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right) && (14)
\end{aligned}$$

Which is maximized when

$$\frac{4\pi(2i+d)}{N} = 0, 2\pi \qquad (15)$$

So,

$$i = \frac{N-d}{2} \qquad (16)$$

if that i exists. There, we have

$$\varepsilon_1^2 = 2A^2\sin^2\left(\frac{2\pi}{N}\right)\left(d+\frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right) \qquad (17)$$

$$\varepsilon_1 = \sqrt{2}A\sin\left(\frac{2\pi}{N}\right)\left(d+\frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right)^{1/2} \qquad (18)$$

The H_1 group dies when all points are within ε_2 of each other.

$$\varepsilon_2^2 = \max_{i,j} \quad 2A^2 \sin^2\left(\frac{2\pi(i-j)}{N}\right)\left(d + \cos\left(\frac{4\pi(i+j+d-1)}{N}\right)\frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right) \tag{19}$$

Perform a change of variables: $x_1 = i - j$, $x_2 = i + j$. Then, $\varepsilon^2(x_1, x_2)$ is separable to two functions, and the maximum is found when both functions are maximized (since they are both everywhere positive).

$$\varepsilon_2^2 = 2A^2 \max_{x_1,x_2} \quad \sin^2\left(\frac{2\pi(x_1)}{N}\right)\left(d + \cos\left(\frac{4\pi(x_2+d-1)}{N}\right)\frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right) \tag{20}$$

And the maximum is achieved when

$$\frac{2\pi x_1}{N} = \pi, \frac{3\pi}{2} \tag{21}$$

$$\frac{4\pi(x_2+d-1)}{N} = 0, \pi \tag{22}$$

if there exist points which satisfy those equations. There, we get

$$\varepsilon_2^2 = 2A^2\left(d + \frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right) \tag{23}$$

$$\varepsilon_2 = \sqrt{2}A\left(d + \frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right)^{1/2} \tag{24}$$

The feature persistence is

$$\varepsilon_2 - \varepsilon_1 = \sqrt{2}A\left(1 - \sin\left(\frac{2\pi}{N}\right)\right)\left(d + \frac{\sin\left(\frac{4\pi d}{N}\right)}{\sin\left(\frac{4\pi}{N}\right)}\right)^{1/2} \tag{25}$$

3 d-Periodic Signals

Consider a signal x which is periodic over the embedding dimension d ($d = N$). In this case, a Fourier transform of the pointwise distance of the embedding to produce a clean simplification.

$$\varepsilon_{i,j} = \|x_{i:i+d} - x_{j:i+d}\|_2^2 \tag{26}$$

Where the indexing is understood modularly.

$$= \sum_{k=0}^{d-1}(x_{i+k} - x_{j+k})^2 \tag{27}$$

Without loss of generality, let $j > i$ and say $\delta = j - i$.

$$= \sum_{k=0}^{d-1}(x_{i+k} - x_{i+k+\delta})^2 \tag{28}$$

Perform a Fourier transform on the inner product. Importantly, the Fourier transform is a unitary transform so distance is preserved. Say $\omega = e^{-2\pi i/d}$ (i here being the imaginary unit). At (30), we rearrange the summation indices.

$$=\frac{1}{d}\sum_{j=0}^{d-1}\left|\sum_{k=0}^{d-1}\omega^{jk}\left(x_{i+k}-x_{i+k+\delta}\right)\right|^2 \tag{29}$$

$$=\frac{1}{d}\sum_{j=0}^{d-1}\left|\sum_{k=0}^{d-1}\omega^{jk}x_{i+k}-\omega^{j(k-\delta)}x_{i+k}\right|^2 \tag{30}$$

$$=\frac{1}{d}\sum_{j=0}^{d-1}\left|\sum_{k=0}^{d-1}\left(\omega^{jk}-\omega^{j(k-\delta)}\right)x_{i+k}\right|^2 \tag{31}$$

$$=\frac{1}{d}\sum_{j=0}^{d-1}\left|\left(1-\omega^{-j\delta}\right)\sum_{k=0}^{d-1}\omega^{jk}x_{i+k}\right|^2 \tag{32}$$

$$=\frac{1}{d}\sum_{j=0}^{d-1}\left(1-\omega^{-j\delta}\right)\overline{\left(1-\omega^{-j\delta}\right)}\left(\sum_{k=0}^{d-1}\omega^{jk}x_{i+k}\right)\overline{\left(\sum_{k=0}^{d-1}\omega^{jk}x_{i+k}\right)} \tag{33}$$

$$=\frac{1}{d}\sum_{j=0}^{d-1}\left(1-\omega^{-j\delta}\right)\left(1-\omega^{j\delta}\right)\left(\sum_{k_1=0}^{d-1}\omega^{jk_1}x_{i+k}\right)\left(\sum_{k_2=0}^{d-1}\omega^{-jk_2}x_{i+k}\right) \tag{34}$$

$$=\frac{1}{d}\sum_{j=0}^{d-1}\left(1-\omega^{-j\delta}-\omega^{j\delta}+1\right)\sum_{k_1=0}^{d-1}\sum_{k_2=0}^{d-1}\omega^{j(k_1-k_2)}x_{i+k_1}x_{i+k_2} \tag{35}$$

$$=\frac{1}{d}\sum_{j=0}^{d-1}\left(2-2\cos\left(\frac{2\pi j\delta}{N}\right)\right)\sum_{k_1=0}^{d-1}\sum_{k_2=0}^{d-1}\omega^{j(k_1-k_2)}x_{i+k_1}x_{i+k_2} \tag{36}$$

We can identify this as an inner product of the Fourier transform of $x_{i:i+d}$ with itself weighted by the diagonal matrix given by $S_{jj}=2-2\cos\left(\frac{2\pi j\delta}{N}\right)$.

$$=\left(\mathcal{F}x_{i:i+d}\right)^H S\left(\mathcal{F}x_{i:i+d}\right) \tag{37}$$

$$=x_{i:i+d}^T\mathcal{F}^H S\mathcal{F}x_{i:i+d} \tag{38}$$

4 Other metrics