Optimal Transport Based Change Point Detection and Time Series Clustering

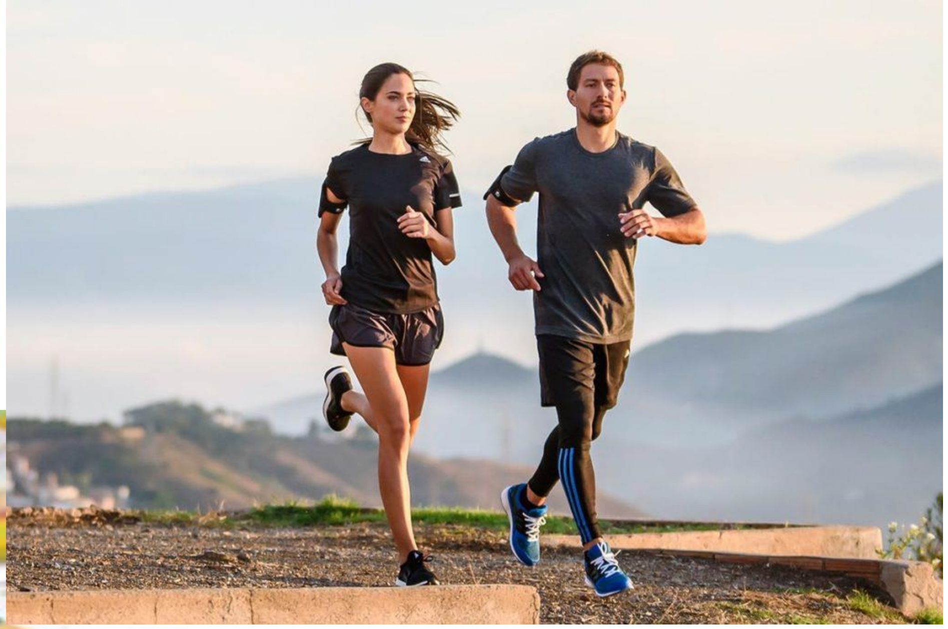
Kevin C. Cheng, Shuchin Aeron, Michael C. Hughes, Erika Hussey, Eric L. Miller

ICASSP 2020

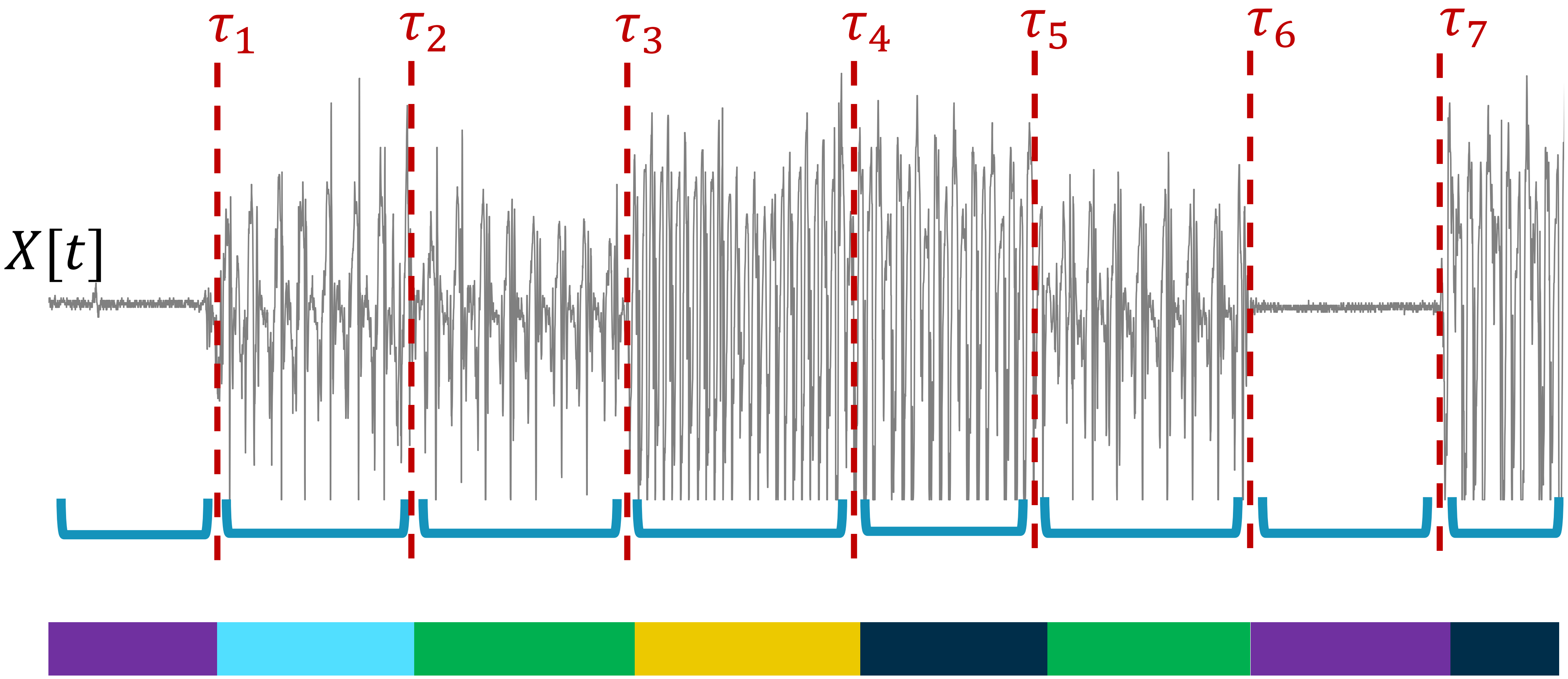
kevin.cheng@tufts.edu

CAREER Award
1553075


40M
20M
0
Q1
Q2
Q3
Q4



Problem Statement



$K = 5$

Overview

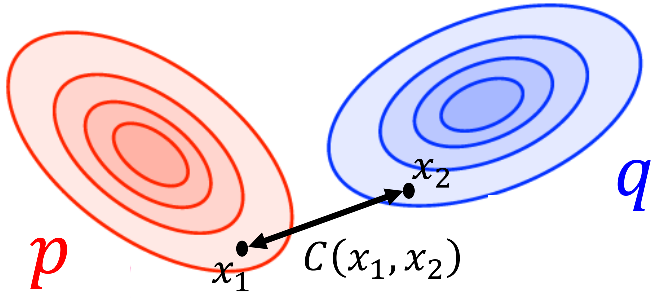
We propose an algorithm for change point detection and time series clustering using optimal transport distances between empirical distribution functions as the measure of similarity.

- Optimal transport benefits
  - Can be computed from empirical samples, no need for density estimation
  - Converges even when distribution supports do not coincide
  - Has a closed form solution in $\mathbb{R}$

G. Peyré and M. Cuturi, “Computational Optimal Transport,” Mar. 2018

Optimal Transport

Continuous Distribution



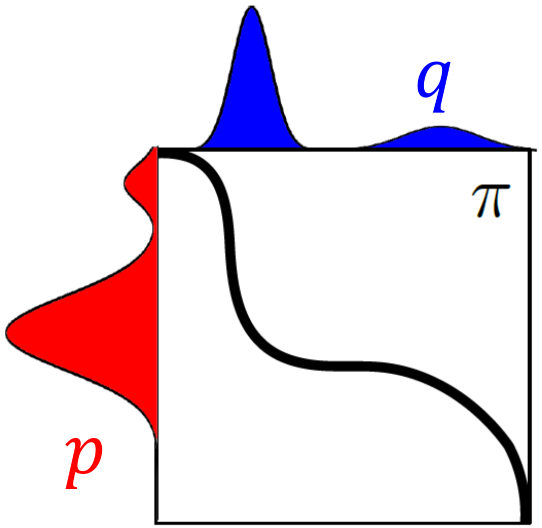
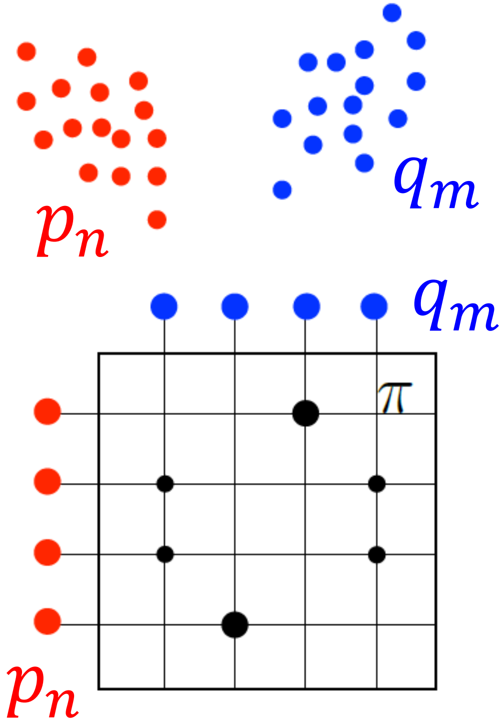


Image credit: [Peyre, Cuturi, 2018]

$$W_2(p,q)^2 = \min_{\pi\in\Pi} \int_R \int_R \|x-y\|_2^2 \pi(x,y)\,dx\,dy$$

$$\int_R \pi(x,y)dx = p\,, \qquad \int_R \pi(x,y)dy = q$$

Discrete Samples



$$W_2(p_n,q_m)^2 = \min_{\pi\in\Pi} \sum_{i=1:n} \sum_{j=1:m} \|p_i - q_j\|_2^2 \pi(i,j)$$

$$p_n = \{p_i\}_{i=1:n}\,, \qquad q_m = \{q_j\}_{j=1:m}$$

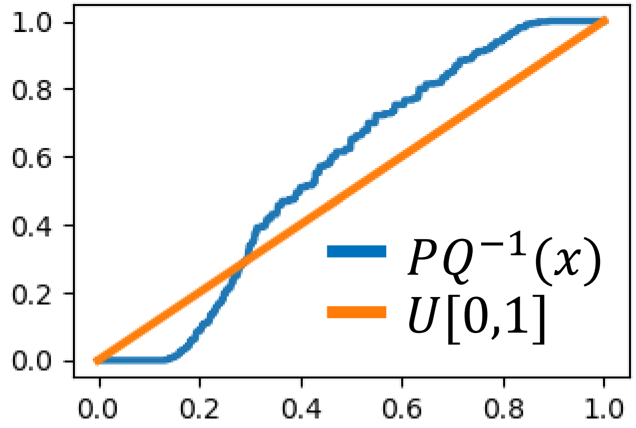
Wasserstein Quantile Test

CDF: $P(x)\colon \mathbb{R} \to [0,1]$

Quantile Function: $P^{-1}(x)\colon [0,1] \to \mathbb{R}$

Q-Q Function: $PQ^{-1}(x)\colon [0,1] \to [0,1]$

$$W_Q(P_n, Q_m) = W_2(P_n Q_m^{-1}, U[0,1])$$

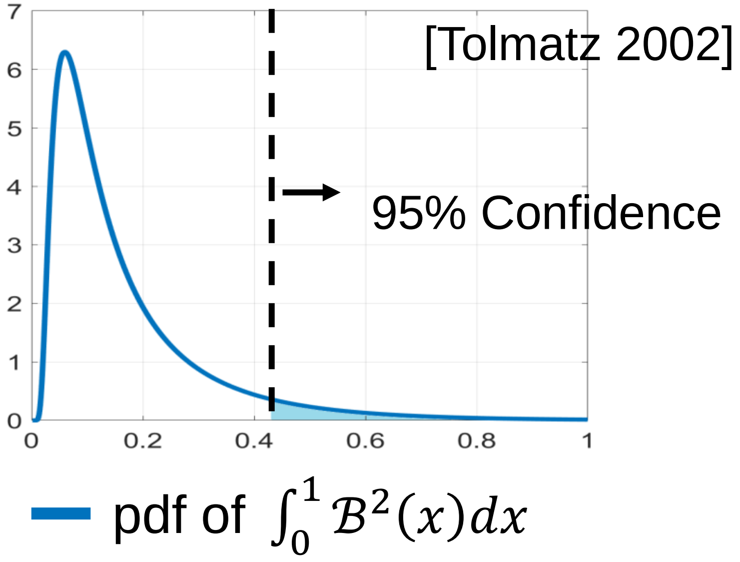


Under the null hypothesis that $P = Q$, asymptotically,

$$\boxed{\frac{nm}{n+m} W_Q(P_m, Q_n) \to_d \int_0^1 \mathcal{B}^2(x)dx}$$

independent of distribution P,

[Ramdas 2015]



$\mathcal{B}(x)$: Brownian bridge stochastic process on the interval $[0,1]$

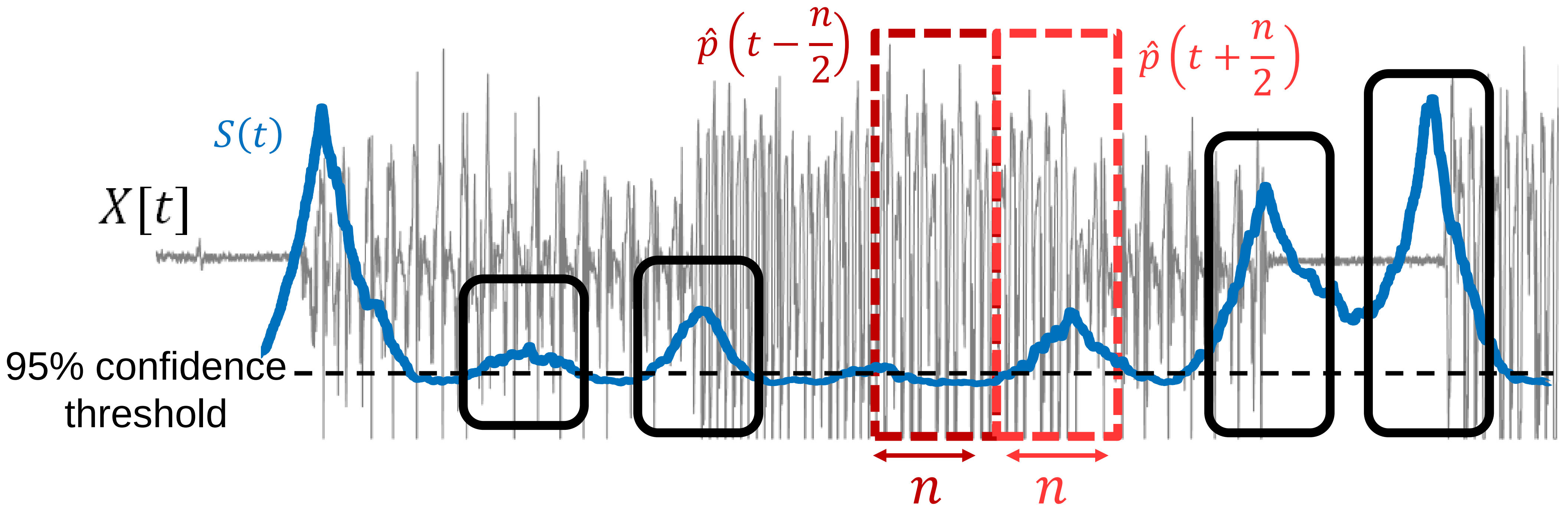
Algorithm Overview

1. Change Point Statistic
2. Matched Filter
3. Clustering of Partitions

Change Point Statistic

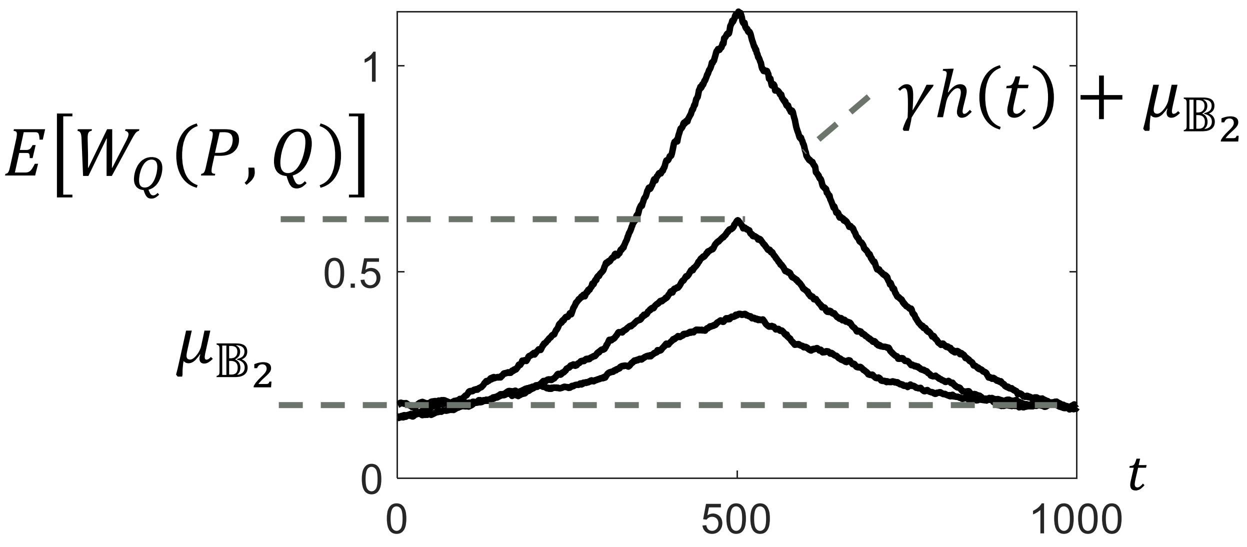
Distribution estimate $\hat{p}(t) = \frac{1}{n}\sum_{i=-n/2}^{n/2-1} \delta(x - X[t-i])$

Change point statistic: $S(t) = \frac{n}{2} W_Q\left(\hat{p}\left(t - \frac{n}{2}\right), \hat{p}\left(t + \frac{n}{2}\right)\right)$

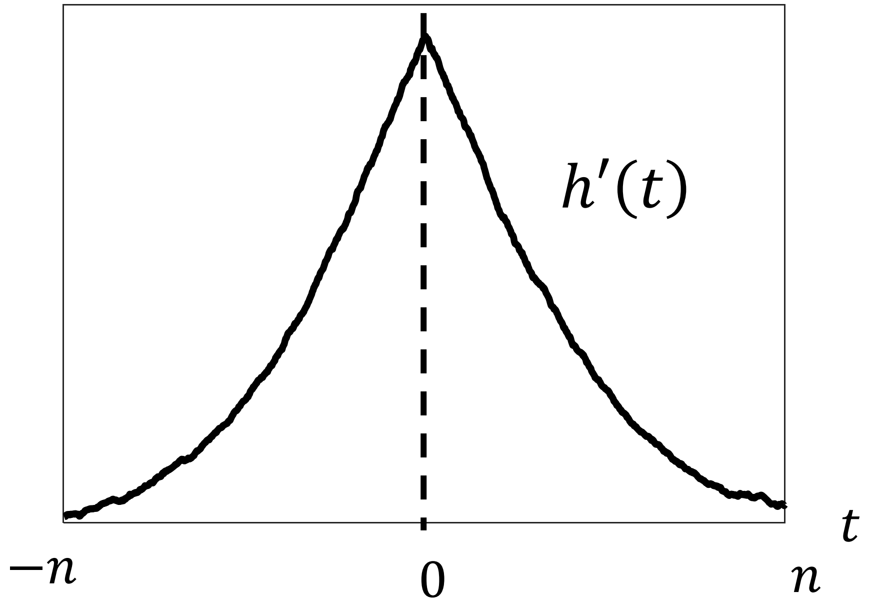


Matched Filter

Results from simulated data



Empirically derived matched filter

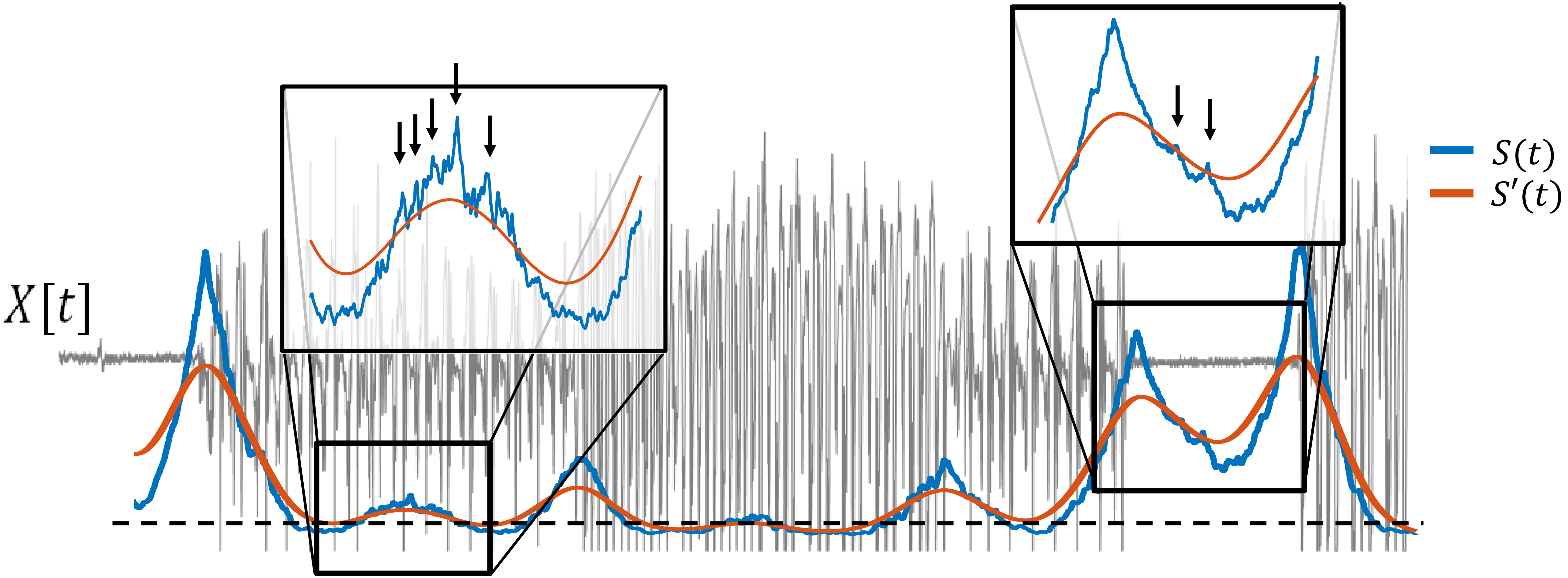


$$h(t) = \alpha \left(\frac{|t-n|}{n} \right)^2$$

[Cheng 2020]

$$\boxed{S'(t) = S(t) * h'(t)}$$

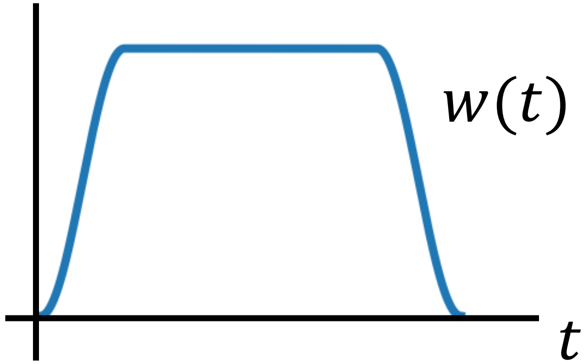
1. Change Point Statistic
2. Matched Filter
3. Clustering of Partitions

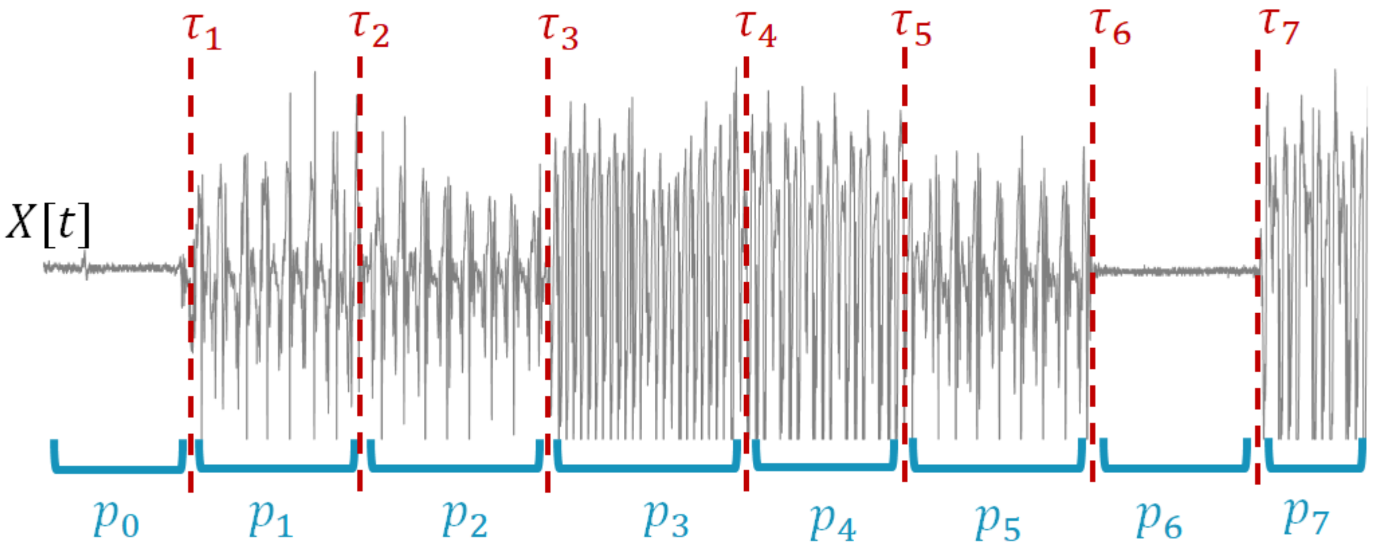


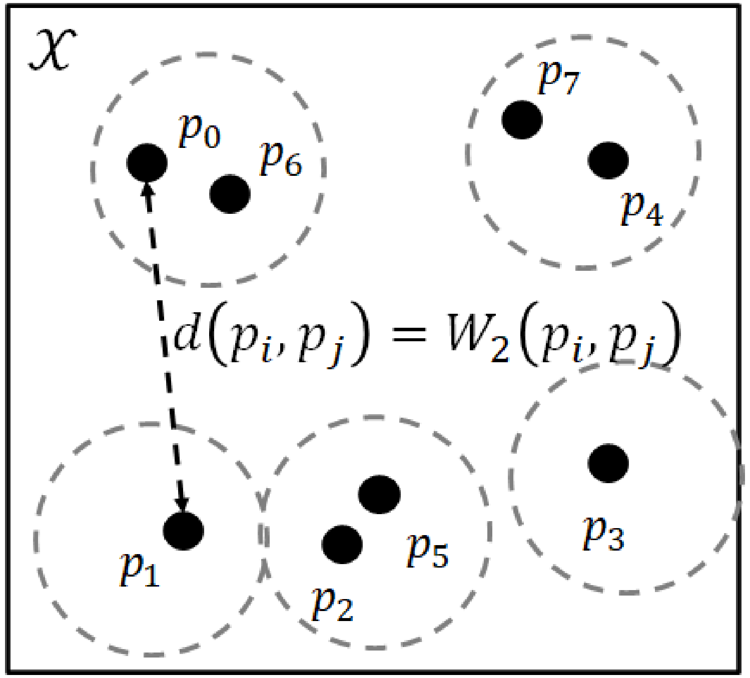
Wasserstein Spectral Clustering

- Distribution of a segment is estimated from weighted data between change points

$$p_i = \frac{1}{\sum_j w_j} \sum_{j=\tau_i}^{\tau_{i+1}} w_j \delta(x - X[j])$$



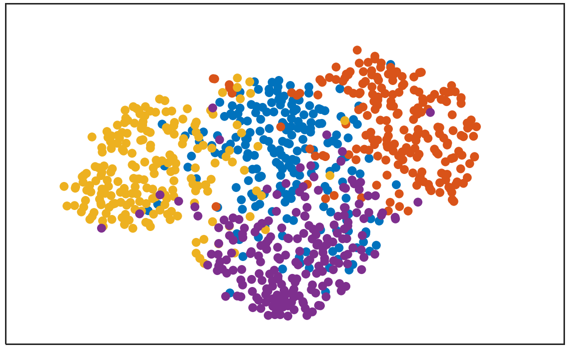




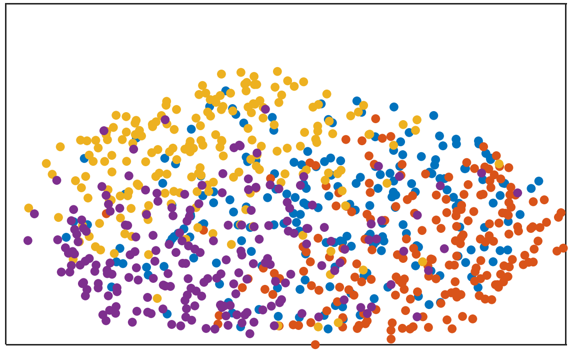
Embedding of Simulated Data

TSNE embedding of 800 sample windows each with 100 samples, from one of 4 distributions

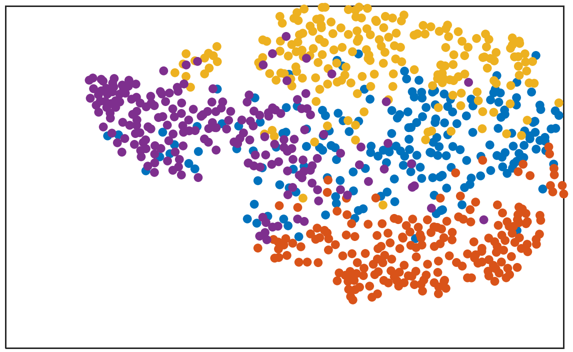
Wasserstein Distance



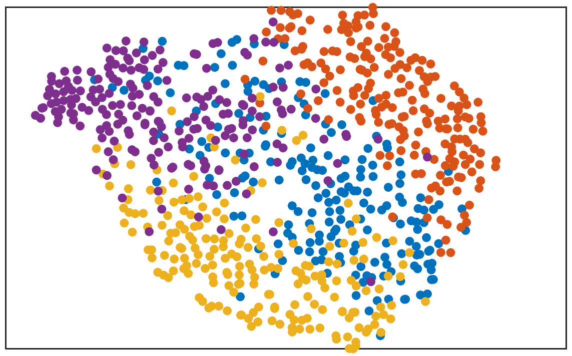
Maximum Mean Discrepancy



Wasserstein Quantile



Kolmogorov-Smirnov

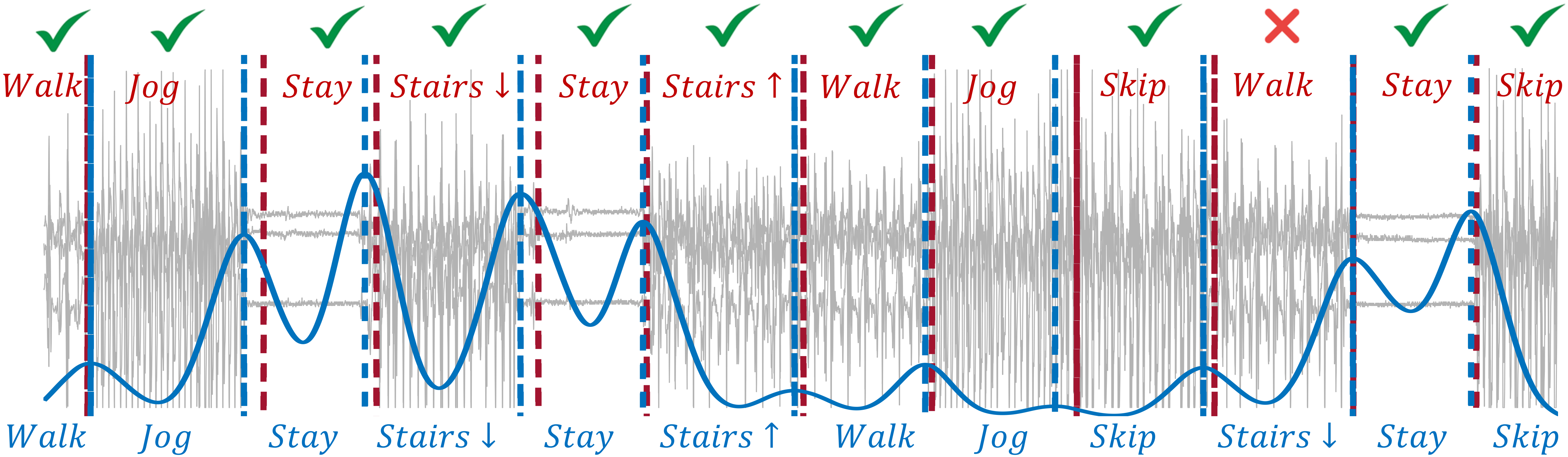


- $Normal(0,1)$
- $Laplace\left(0, \frac{1}{\sqrt{2}}\right)$
- $Normal(0.2,1)$
- $Normal(0,1.2)$

Wasserstein Distance shows the most separability between clusters

Human Activity Sensing (HASCPAC-2016)

Ground Truth



Computed Change Point and Segment Cluster Assignments

71% Precision, 79% Recall, 66% Label Accuracy

Results Summary

Change Point Detection

Dataset	K	n	Change Point F1-Score	
			W-Quantile	MMD
HASCPAC-2016	6	500	0.748	0.713
Beedance	3	14	0.647	0.625
HASC2011	6	500	0.824	0.770
ECG200	2	100	0.637	0.583

Time Series Clustering

Label Accuracy	
Ground Truth	W-Quantile
0.789	0.658
0.705	0.651
0.565	0.398
0.864	0.708

Change point detection using the Wasserstein quantile shows improved results over MMD,
Time series clustering given ground truth change points comparable to other unsupervised methods.

Conclusions

- We propose a method for unsupervised optimal transport based change point detection and time series segment clustering.
- We use the Wasserstein quantile test for change point detection as a distribution-independent statistical test with a empirically derived matched filter.
- Time series clustering through spectral clustering using the Wasserstein distance as the similarity measure between empirical distributions.

Special Thanks To:

Eric L. Miller

Tufts UNIVERSITY

Shuchin Aeron

Tufts UNIVERSITY

NSF CAREER Award 1553075

Michael C. Hughes

Tufts UNIVERSITY

Erika Hussey

DEVCOM ARMAMENTS CENTER

CABCS CENTER FOR APPLIED BRAIN & COGNITIVE SCIENCES

Questions?

kevin.cheng@tufts.edu

Paper

https://tinyurl.com/OtCpdTsc

Contact

Code

https://tinyurl.com/OtCpdTSC-Code