Fast Global Alignment Kernels

Time Series

05.05.2015

Rodrigo Frassetto Nogueira
Thanos Papadopoulos

NYU

Agenda

- Motivation
- Dynamic Time Warping (DTW)
- Global Alignment (GA) Kernels
- Experiments
- Conclusion

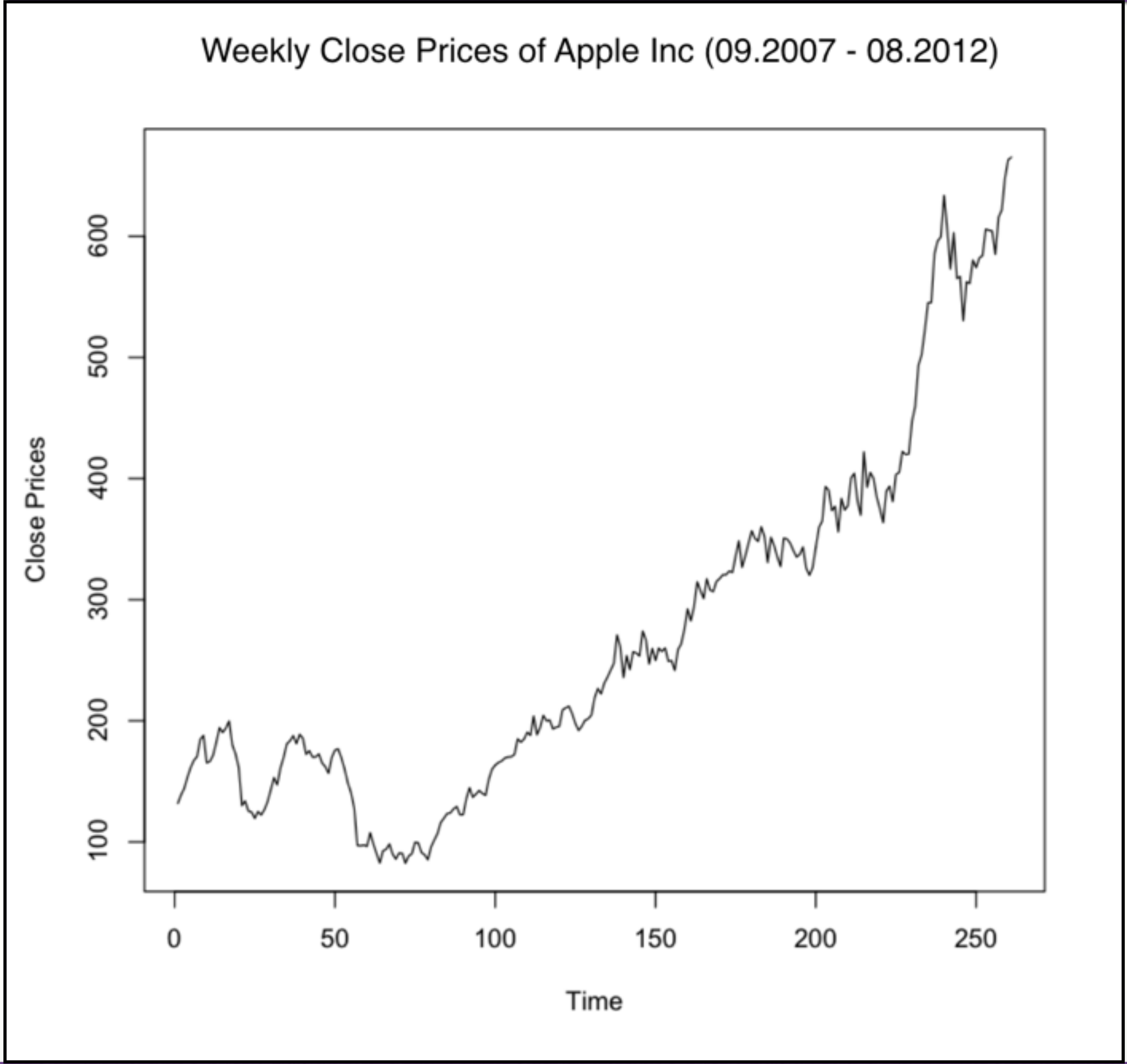
Agenda

- **Motivation**
  - **Time series Introduction**
  - Problem Formulation
- Dynamic Time Warping (DTW)
- Global Alignment (GA) Kernels
- Experiments
- Conclusion

Example #1

- Stock Price Forecasting

  Apple Inc (9.2007 - 8.2012)



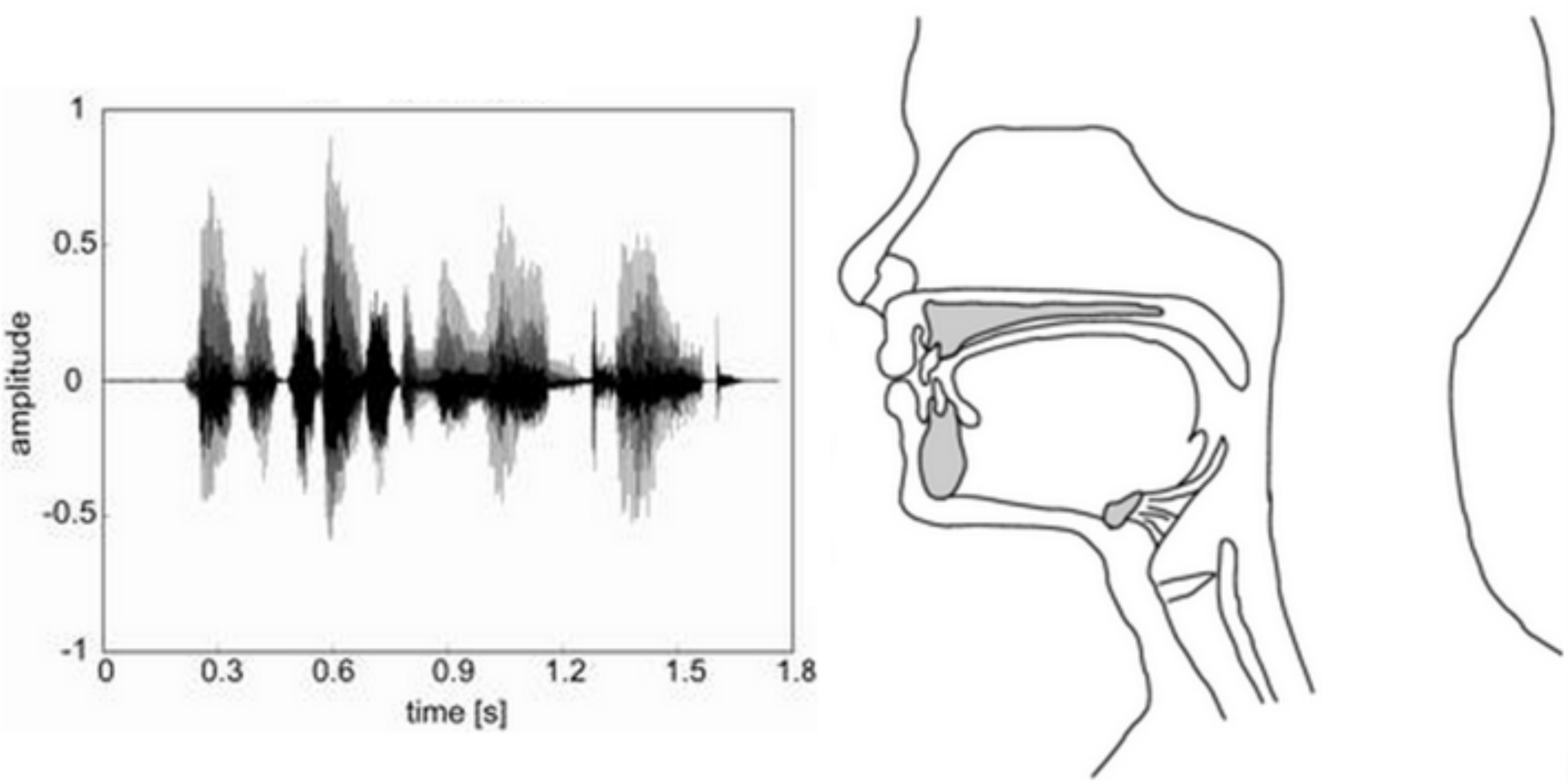
Weekly Close Prices of Apple Inc (09.2007 - 08.2012)
Close Prices
100
200
300
400
500
600
0
50
100
150
200
250
Time

Example #2

- Caltrans Performance Measurement System (PeMS)
  - occupancy rate in San Francisco bay area freeways - [0,1]
  - 963 sensors (different car lanes)
  - data every 10’
  - time series per day: dimension - 963, length - 6*24=144

Example #3

- Speech signals



Agenda

- **Motivation**
  - Time series Introduction
  - **Problem Formulation**
- Dynamic Time Warping (DTW)
- Global Alignment (GA) Kernels
- Experiments
- Conclusion

Problem Formulation

- Goal: Find adequate kernels for time series
  - handle variable length
  - be positive definite
  - low computational cost

How we start?

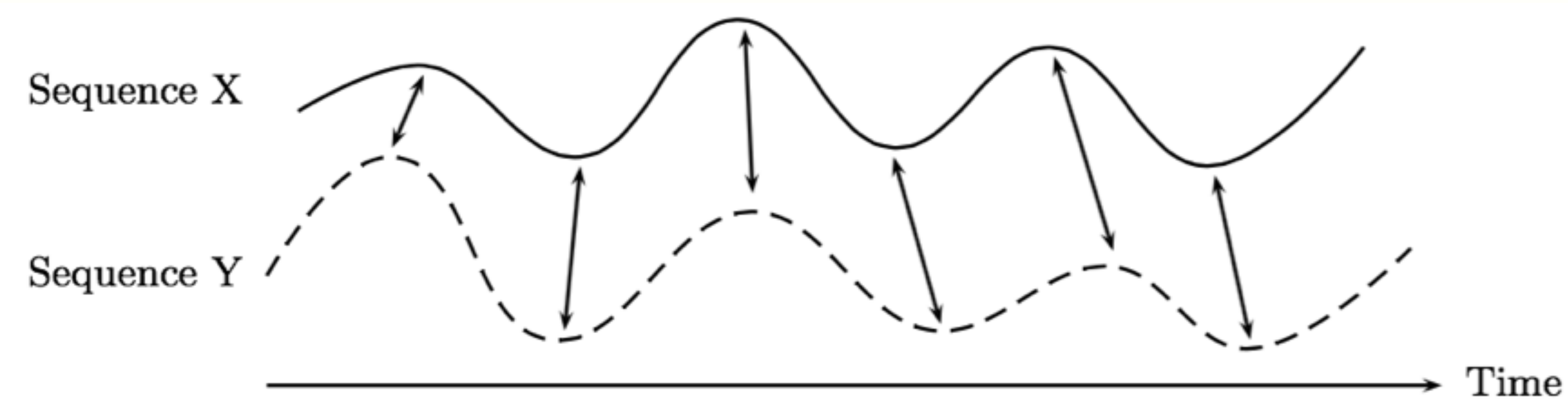
Similarity Measure: $K(x, y) = \langle \phi(x), \phi(y) \rangle$

Agenda

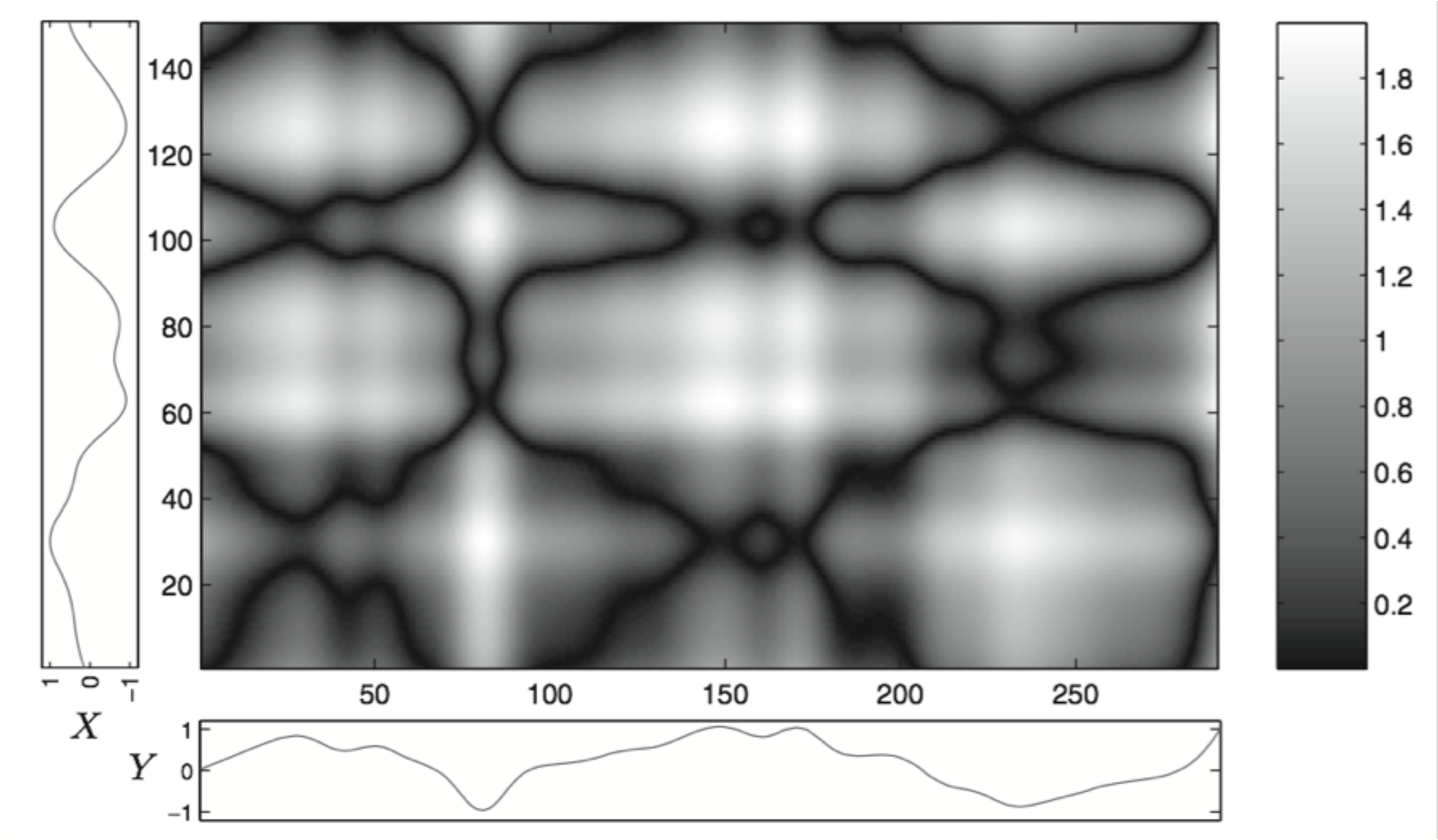
- Motivation
- **Dynamic Time Warping (DTW)**
- Global Alignment (GA) Kernels
- Experiments
- Conclusion

Dynamic Time Warping

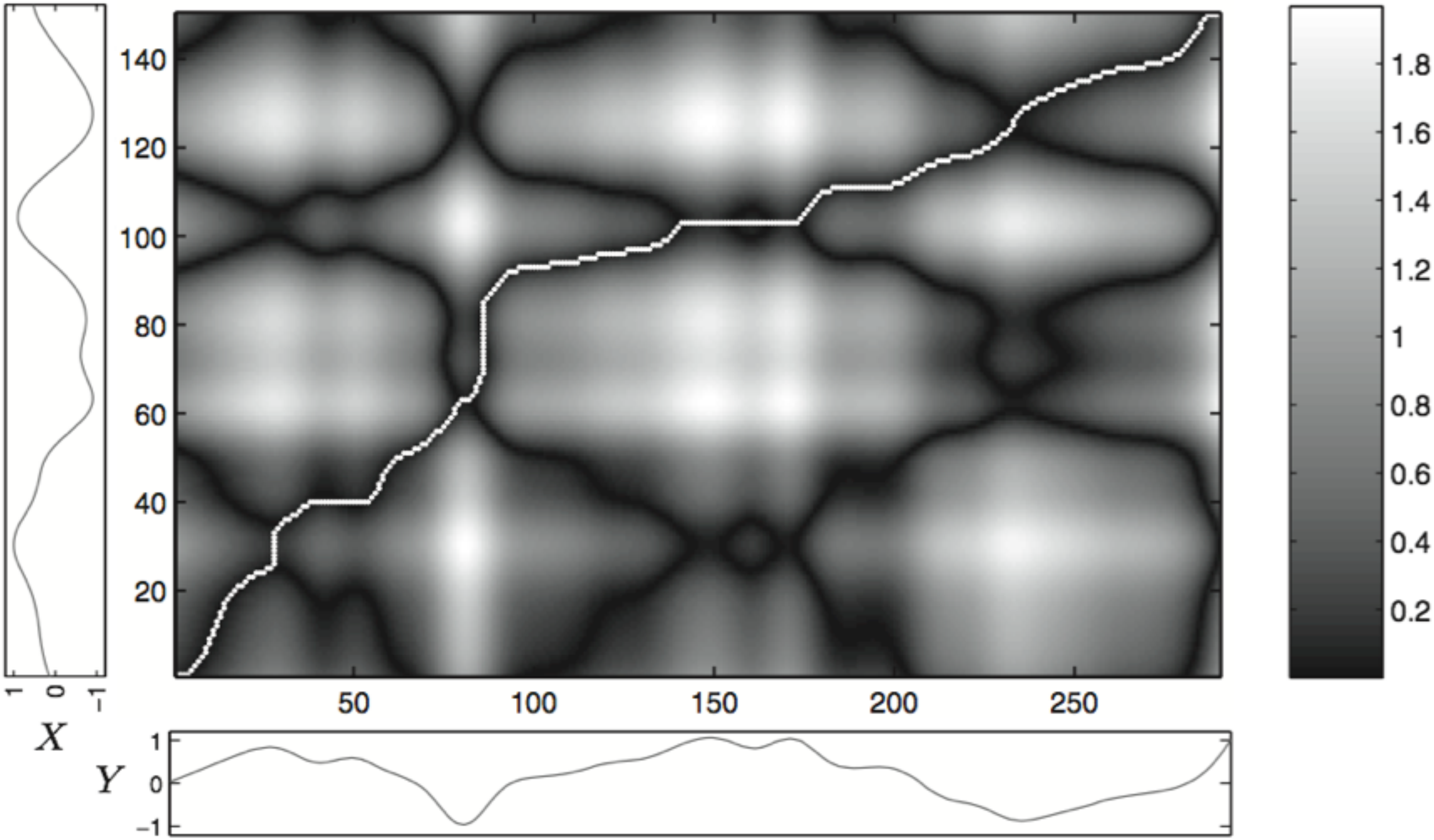
- pairwise comparisons?
- alignment:
  associate each element of sequence *X* to one or more elements of sequence *Y* and vice-versa



DTW Cost Matrix

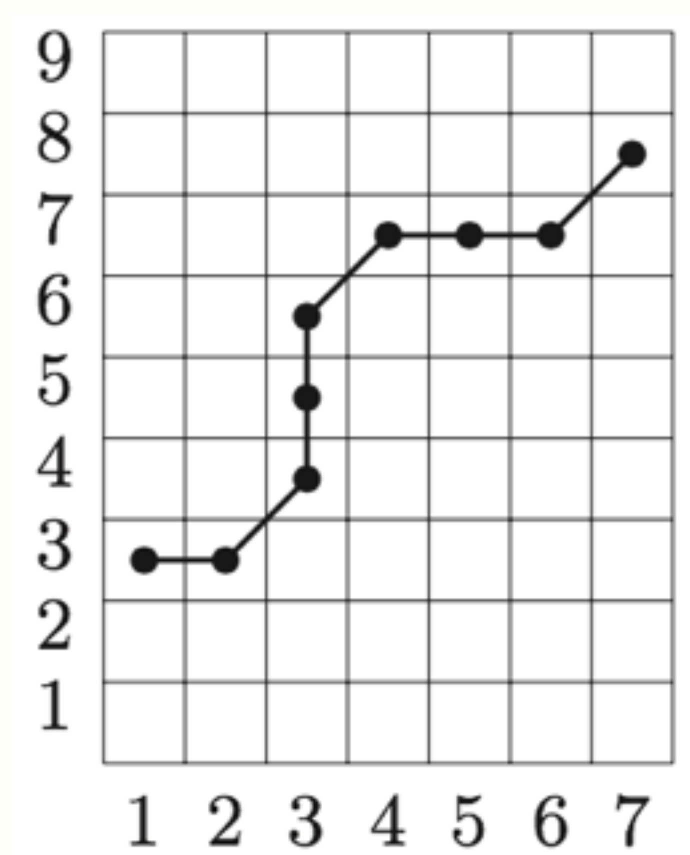


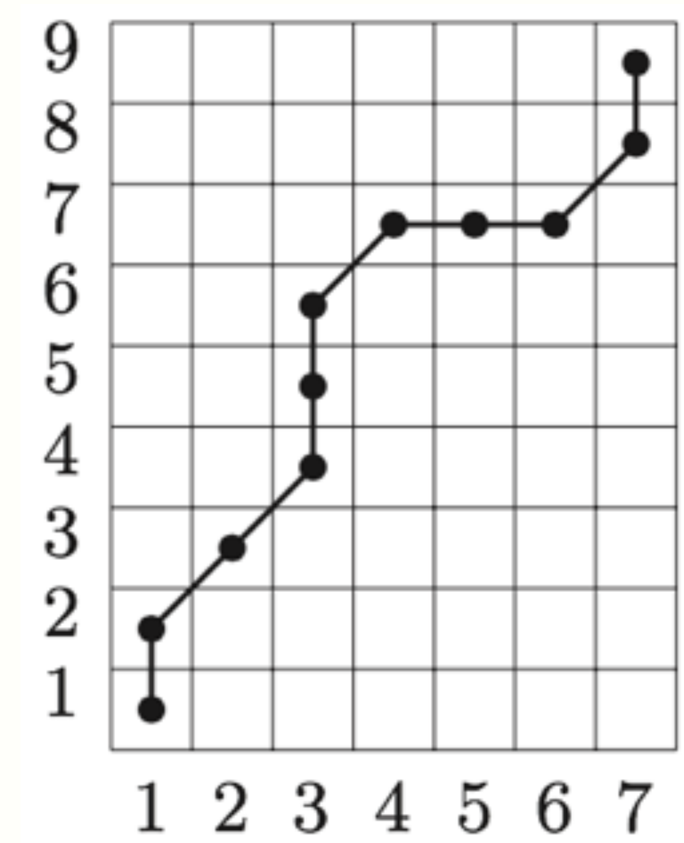
Optimal Alignment



DTW definition

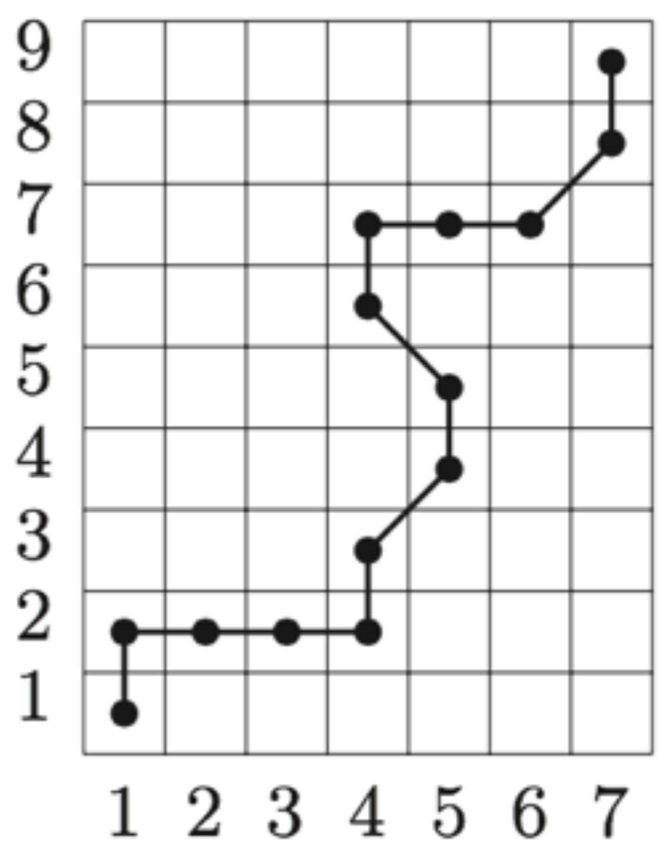
- warping functions: $\pi_1(i), \pi_2(j)$

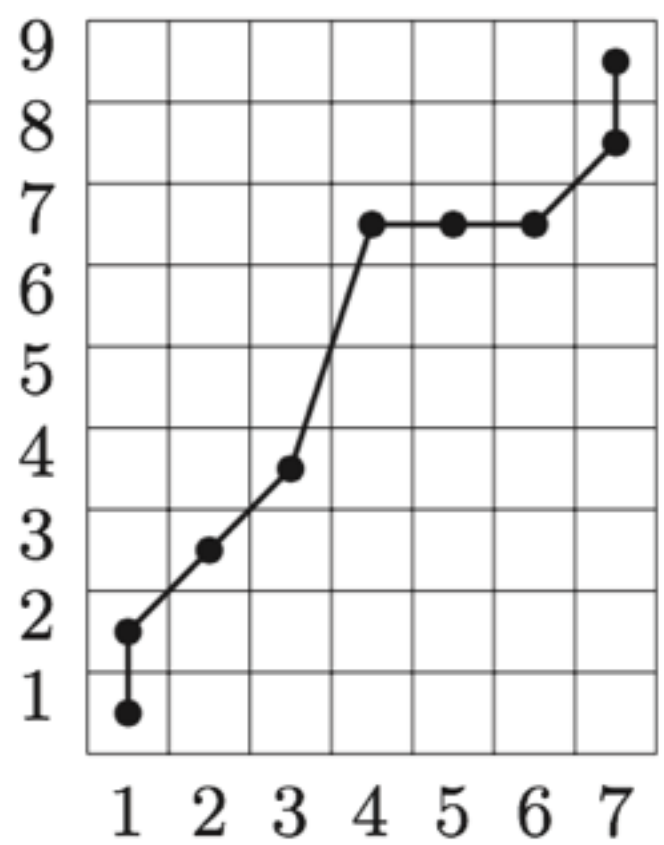


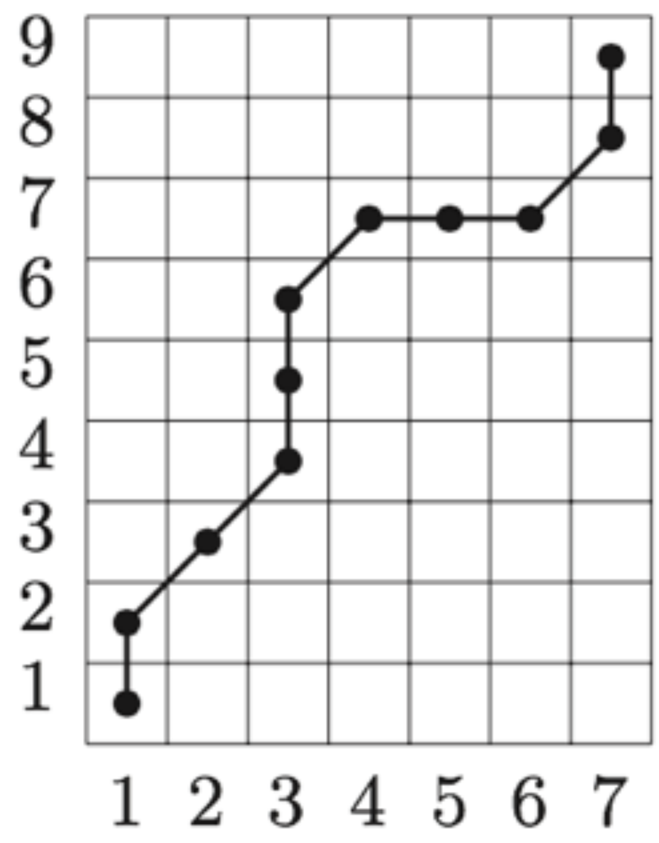


DTW definition

- warping functions: $\pi_1(i), \pi_2(j)$
- moves: $\rightarrow, \uparrow, \nearrow$







DTW definition

- warping functions: $\pi_1(i), \pi_2(j)$

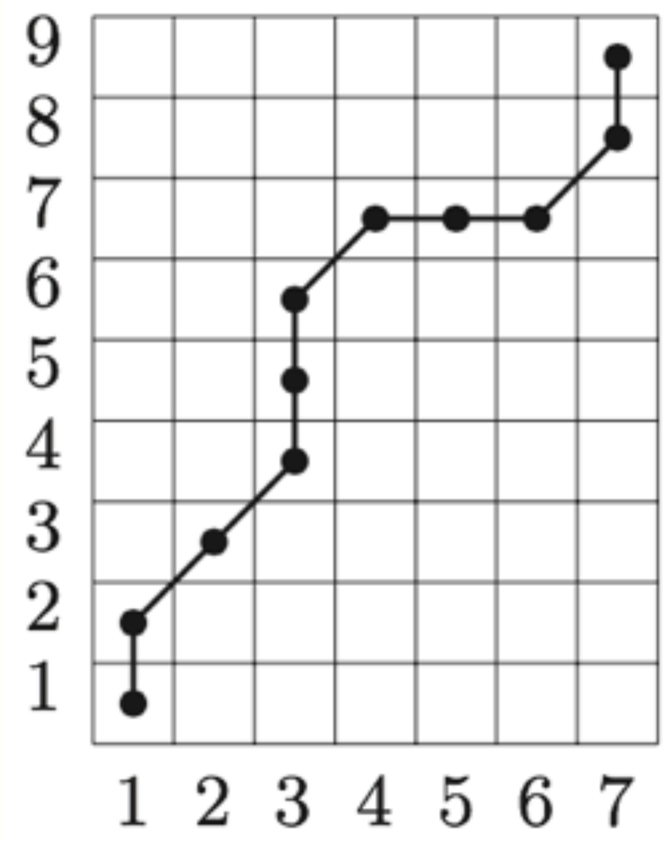
- moves: $\rightarrow, \uparrow, \nearrow$

- cost per alignment π:

$$D_{x,y}(\pi) = \sum_{i=1}^{|\pi|} \varphi(x_{\pi_1(i)}, y_{\pi_2(i)})$$

- optimal alignment:

$$DTW(x,y) = \min_{\pi \in A(n,m)} D_{x,y}(\pi)$$



Why not DTW?

- not PDS
- high computational cost, O(dnm)

Agenda

- Motivation
- Dynamic Time Warping (DTW)
- **Global Alignment (GA) Kernels**
    - **Definition**
    - Diagonal Dominance
    - Positive Definiteness
    - Computational Cost
- Experiments
- Conclusion

GA kernels

- soft maximum motivation, $log(\sum_i e^{x_i})$:

$$k_{GA} = \sum_{\pi \in A(n,m)} e^{-D_{x,y}(\pi)} = \sum_{\pi \in A(n,m)} e^{-\sum_{i=1}^{|\pi|} \varphi(x_{\pi_1(i)}, y_{\pi_2(i)})}$$

- by defining $\kappa = e^{-\varphi}$:

$$k_{GA} = \sum_{\pi \in A(n,m)} \prod_{i=1}^{|\pi|} \kappa(x_{\pi_1(i)}, y_{\pi_2(i)})$$

- whole spectrum of costs/alignments

Agenda

- Motivation
- Dynamic Time Warping (DTW)
- **Global Alignment (GA) Kernels**
    - Definition
    - **Diagonal Dominance**
    - Positive Definiteness
    - Computational Cost
- Experiments
- Conclusion

Diagonal Dominance

- off-diagonal entries far smaller than trace (Gram matrix)

- Solution:

$$\kappa = e^{-\lambda \varphi(x_{\pi_1(i)}, y_{\pi_2(i)})}, \quad \lambda > 0$$

Behavior in the limits of λ

- Let $\{X_1, X_2, ..., X_p\}$ be a sample of time series

$$k_{GA}(x,y) = \sum_{\pi \in A(n,m)} \prod_{i=1}^{|\pi|} e^{-\lambda \cdot \varphi(x_{\pi_1(i)}, y_{\pi_2(i)})}$$

1. All samples have same length n
   - When $\lambda \to \infty, \quad k_{GA} \to I_p$
   - When $\lambda \to 0, \quad k_{GA} \to D(n,n) \cdot 1_{p,p}$

Delannoy numbers

1	1	1	1	1	1	1	1	1	⋯
1	3	5	7	9	11	13	15	17	⋯
1	5	13	25	41	61	85	113	145	⋯
1	7	25	63	129	231	377	575	833	⋯
1	9	41	129	321	681	1289	2241	3649	⋯
1	11	61	231	681	1683	3653	7183	13073	⋯

$$D(n,m) = D(n,m-1) + D(n-1,m) + D(n-1,m-1)$$

Behavior in the limits of λ

- Let $\{X_1, X_2, ..., X_p\}$ be a sample of time series

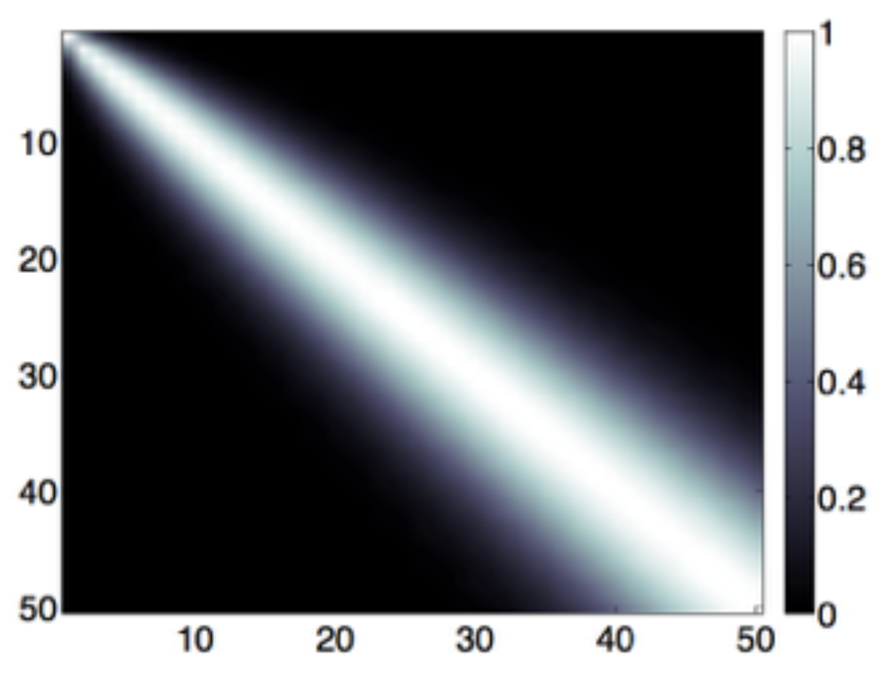
$$k_{GA}(x, y) = \sum_{\pi \in A(n,m)} \prod_{i=1}^{|\pi|} e^{-\lambda \cdot \varphi(x_{\pi_1(i)}, y_{\pi_2(i)})}$$

2. Samples have different lengths

   - When $\lambda \to \infty, \quad k_{GA} \to I_p$
   - When $\lambda \to 0, \quad k_{GA}(X_i, X_j) \to D(|X_i|, |X_j|)$

Diagonal Dominance

- problem arises when length varies and λ tends to 0
- n sequences with length 1 to n



- Empirically: for $\lambda \approx 0 \Rightarrow \frac{1}{2} \leq \frac{n}{m} \leq 2$

Agenda

- Motivation
- Dynamic Time Warping (DTW)
- **Global Alignment (GA) Kernels**
    - Definition
    - Diagonal Dominance
    - **Positive Definiteness**
    - Computational Cost
- Experiments
- Conclusion

Positive Definiteness

- if κ is p.d. what about k_{GA} ?
- mapping kernels: $$k(x, y) = \sum_{(x_i, y_i) \in M(x,y)} \kappa_l(x_i, y_i)$$

κ is a local kernel on substructures of x, y
M is a mapping set

GA kernels as Mapping kernels

we don't know

- if κ is p.d. what about k_{GA} ?
- For GA kernels: $\kappa_l(x_{\pi_1}, y_{\pi_2}) = \prod_{i=1}^{|\pi|} \kappa(x_{\pi_1(i)}, y_{\pi_2(i)})$

$$M_{GA}(x,y) = \{(x_{\pi_1}, y_{\pi_2}) | \pi = (\pi_1, \pi_2) \in A(n,m)\}$$

- Theorem 1: k_{GA} is p.d. if and only if M is transitive

$$(x_i, y_i) \in M(x,y), (y_i, z_i) \in M(y,z) \Rightarrow (x_i, z_i) \in M(x,z)$$

- Lemma 1: M_{GA} is not transitive

Constraints on κ

- Theorem 2: If κ is p.d. and $\frac{\kappa}{1+\kappa}$ is p.d., k_{GA} is p.d.
- Geometric Divisibility (g.d.):
  - mapping $\tau(x) = \frac{x}{1+x} : R_+ \to [0,1)$
  - inverse mapping $\tau^{-1}(x) = \frac{x}{1-x} : [0,1) \to R_+$
  - f is g.d. if τf is p.d.
- Lemma 2: If κ is g.d. then k_{GA} is p.d.

Constraints on κ

- Infinite divisibility (i.d.): κ is i.d. iff $-log(\kappa)$ is n.d.
- Lemma 3: For any i.d. kernel κ s.t. $0 < \kappa < 1$, $\tau^{-1}\kappa$ is g.d. and i.d.
- how to construct a local kernel?
- Example: Gaussian kernel, κ_σ , is i.d., thus $\tau^{-1}(\frac{\kappa_\sigma}{2})$ is i.d. and g.d.
  Also, $\varphi = -log(\tau^{-1}(\frac{\kappa_\sigma}{2}))$ is n.d. and we set $\kappa = e^{-\lambda\varphi}$

Agenda

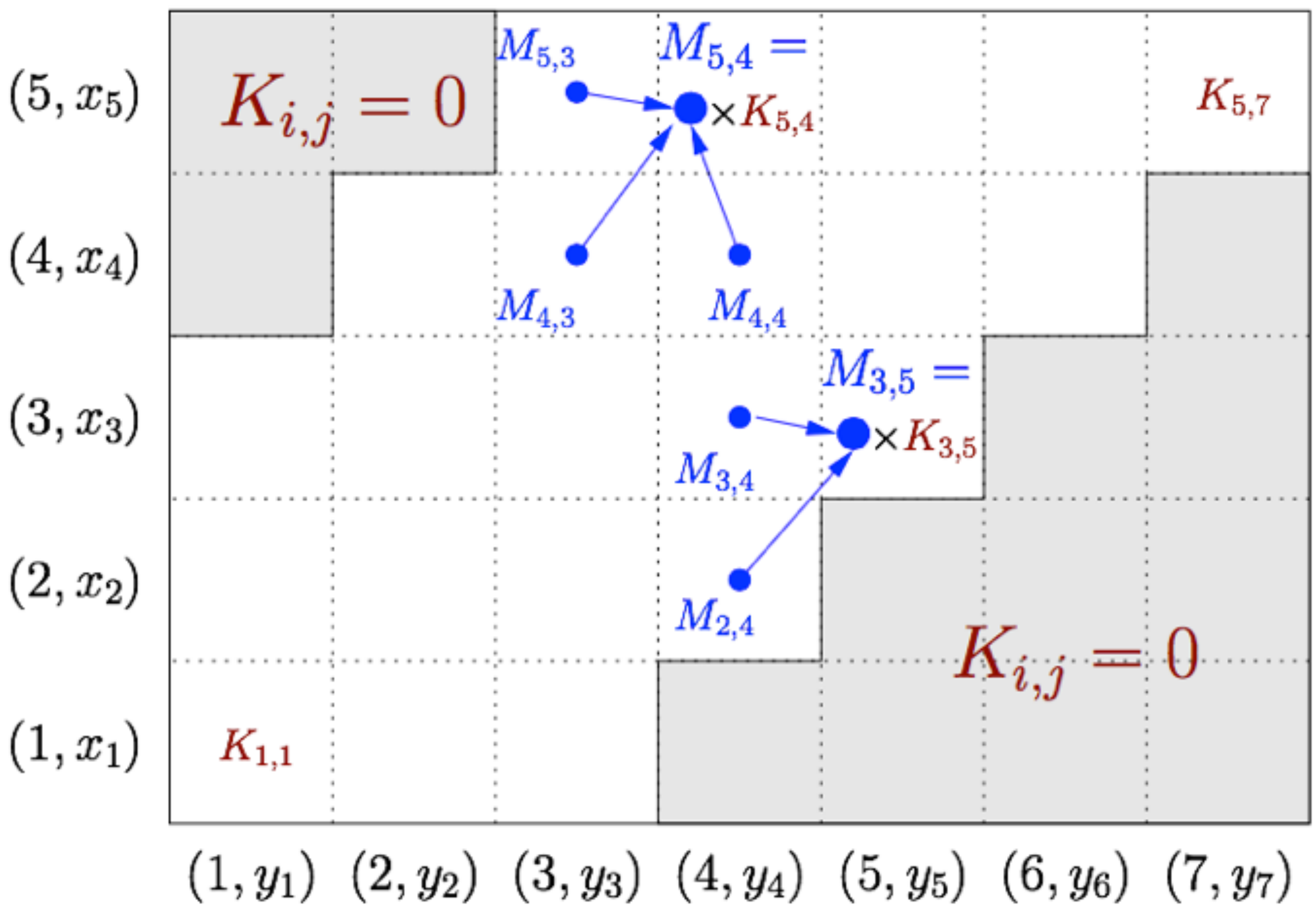
- Motivation
- Dynamic Time Warping (DTW)
- **Global Alignment (GA) Kernels**
    - Definition
    - Diagonal Dominance
    - Positive Definiteness
    - **Computational Cost**
- Experiments
- Conclusion

Time Complexity

- DTW time complexity: O(dnm)
- What to do?
- Ignore "bad" alignments - keep alignments close to the diagonal:

$$D^{\gamma}_{x,y}(\pi) = \sum_{i=1}^{|\pi|} \gamma_{\pi_1(i),\pi_2(i)} \cdot \varphi(x_{\pi_1(i)}, y_{\pi_2(i)})$$

$$\gamma_{i,j} = \begin{cases} 1, & |i-j| < T \\ \infty, & |i-j| \geq T \end{cases}$$



$$M_{i,j} = \kappa(x_i, y_j)(M_{i-1,j-1} + M_{i,j-1} + M_{i-1,j})$$

$$(2T-1) \cdot min(n,m) - T(T-1)/2 \quad \Rightarrow \quad O(2Tmin(n,m))$$

Triangular GA Kernels

- if κ is infinite divisible, and ω is a p.d. kernel in N, then $\tau^{-1}(\omega \otimes \kappa)$ as a local kernel gives a p.d. GA kernel

- common choice for ω:

$$\omega(i,j) = \left(1 - \frac{|i-j|}{T}\right)_+$$

- Example: $\tau^{-1}(\omega \otimes \frac{1}{2}\kappa_\sigma)(i,x;j,y) = \frac{\omega(i,j)\kappa(x,y)}{2-\omega(i,j)\kappa(x,y)}$

Agenda

- Motivation
- Dynamic Time Warping (DTW)
- Global Alignment (GA) Kernels
- **Experiments**
- Conclusion

Kernels to compare

- DTW: $k_{DTW} = e^{-\frac{1}{t}DTW}$
- DTW SC: $k_{SC} = e^{-\frac{1}{t}DTW_{SC}}$

$$DTW_{SC}(x,y) = \min_{\pi \in A(n,m)} D^{\gamma}_{x,y}(\pi)$$

- DTAK: $\left(k_{DTAK}(x,y)\right)^{\frac{1}{t}}$

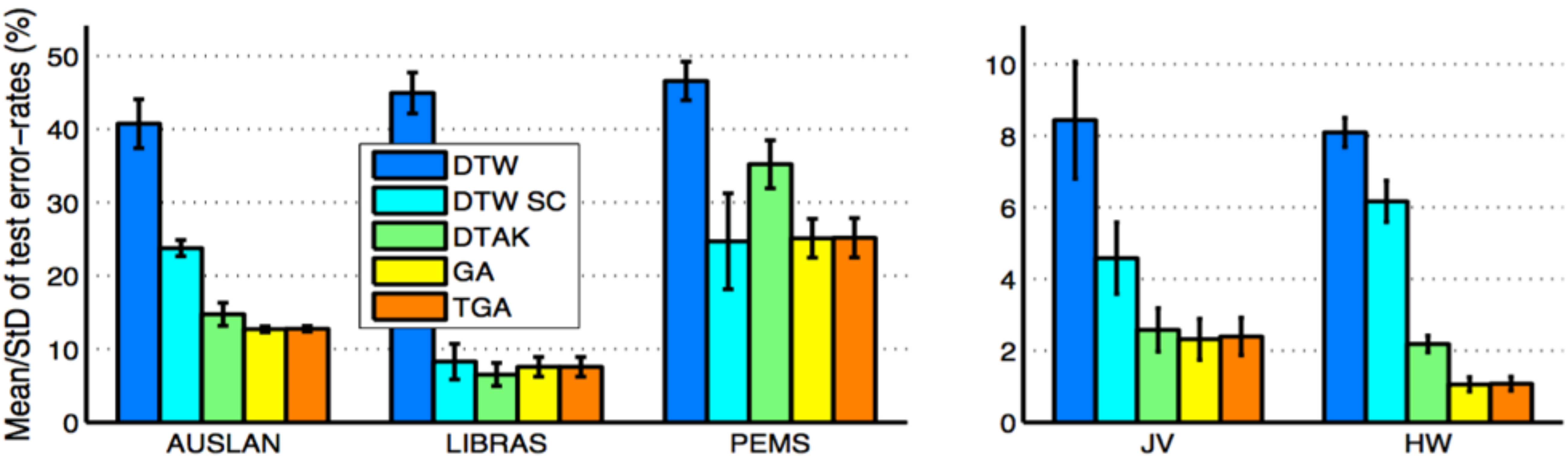
$$k_{DTAK}(x,y) = \max_{\pi \in A(n,m)} \sum_{i=1}^{|\pi|} \kappa_{\sigma}(x_{\pi_1(i)}, y_{\pi_2(i)})$$

- GA: $\kappa = e^{-log\left(\tau^{-1}(\frac{1}{2}\kappa_{\sigma})\right)}$
- TGA: $\kappa = \tau^{-1}(\omega \otimes \frac{1}{2}\kappa_{\sigma})$

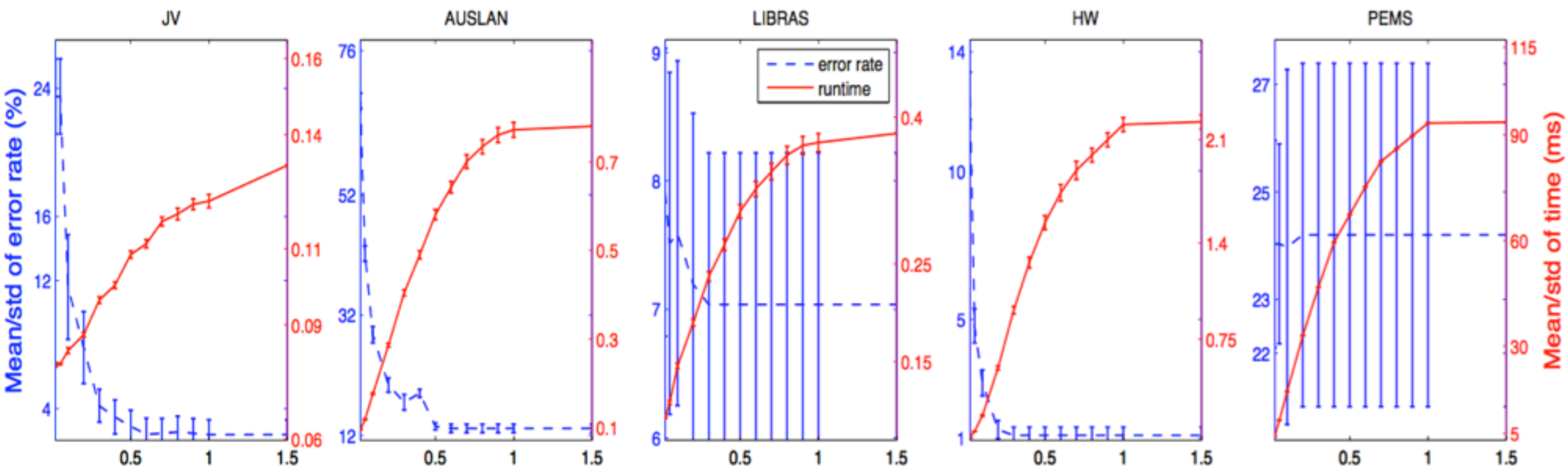
Datasets

Database	d	n range, **med**(n)	**classes**	**# points**
Japanese Vowels	12	7-29, 15	9	640
Libras	2	45	15	945
Handwritten Characters	3	60-182, 122	20	2858
AUSLAN	22	45-136, 55	95	2465
PEMS	963	144	7	440

Classification error rates



Performance and speed



Agenda

- Motivation
- Dynamic Time Warping (DTW)
- Global Alignment (GA) Kernels
- Experiments
- **Conclusion**

Conclusion

$O(T \cdot min(n, m))$

PDS
variable size

Fast Global Alignment Kernels

wide spectrum of alignments

DTW based