



PennState

# A distance for HMMs Based on Aggregated Wasserstein and State Registration

Yukun Chen<sup>1</sup>, Jianbo Ye<sup>1</sup>, Jia Li<sup>2</sup>

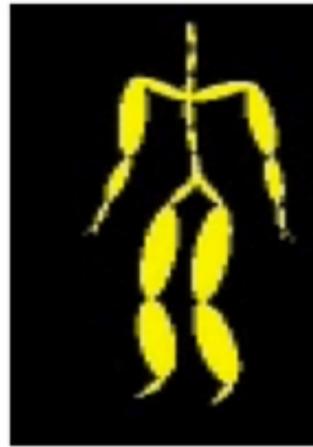
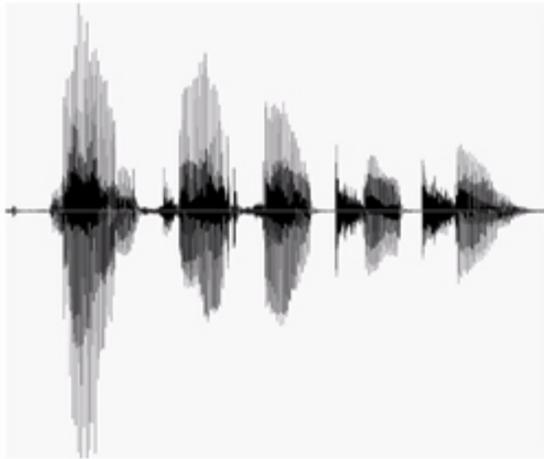
<sup>1</sup>College of Information Sciences & Technology

<sup>2</sup>Department of Statistics

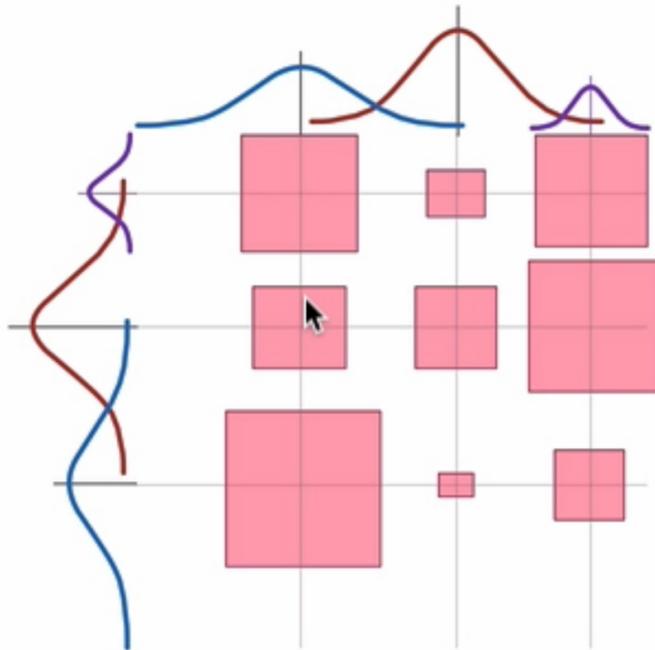
Penn State University

# Hidden Markov Model (HMM)

- A generative model for sequence data/time series.
- Wide Applications



- Key parameters for GMM-HMM:



# Distance for HMMs

- Critical for HMM based sequence **retrieval**, **classification** [ECCV06], **clustering** [NIPS97, JMLR14] etc.
- KL-Based distance [Juang *et al.* 1985]
  - Rely on original sequence data
  - Or need to generate sequence from one model
- Minimized Aggregated Wasserstein (MAW) and Improved Aggregated Wasserstein (IAW)
  - Solely rely on the parameters of HMMs
  - Better differentiation ability

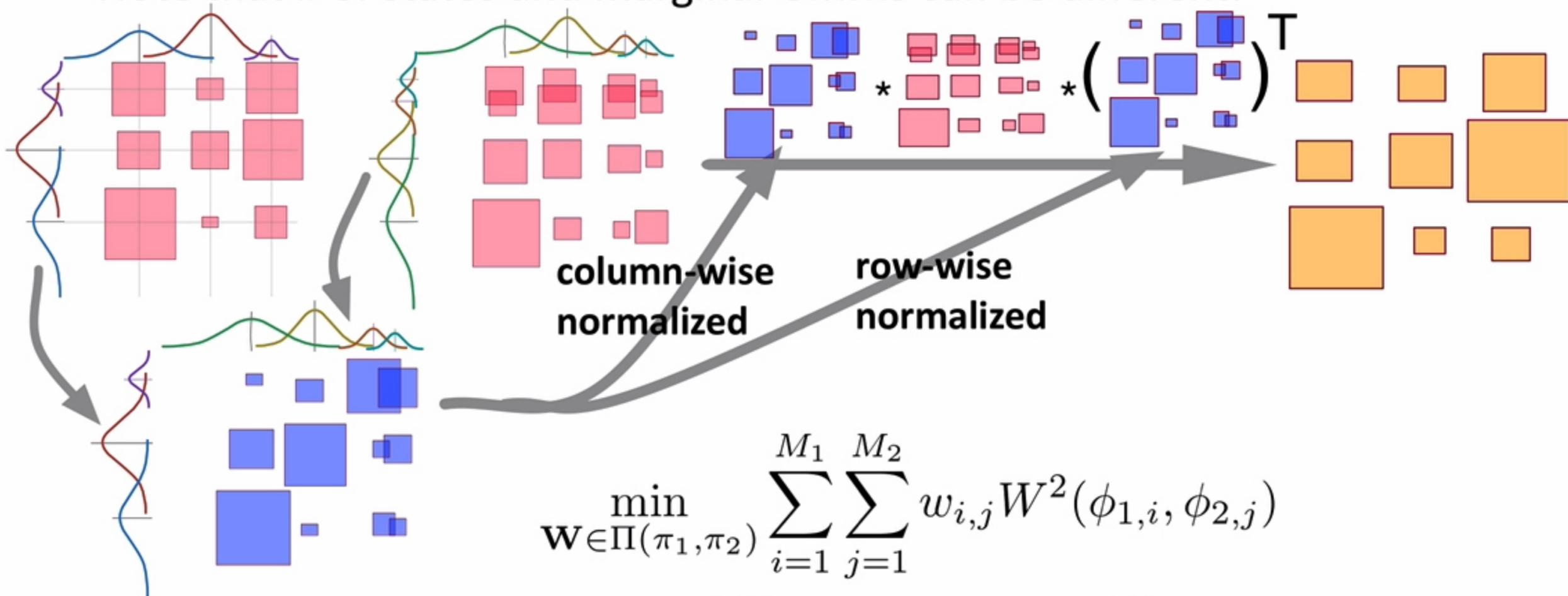
# Minimized Aggregated Wasserstein (MAW)

Step 1:  
State Registration

Step 2:  
Diff. b/w Marginal GMMs

Step 3:  
Diff. b/w Transition Matrices

Note that # of states and marginal GMMs can be different:

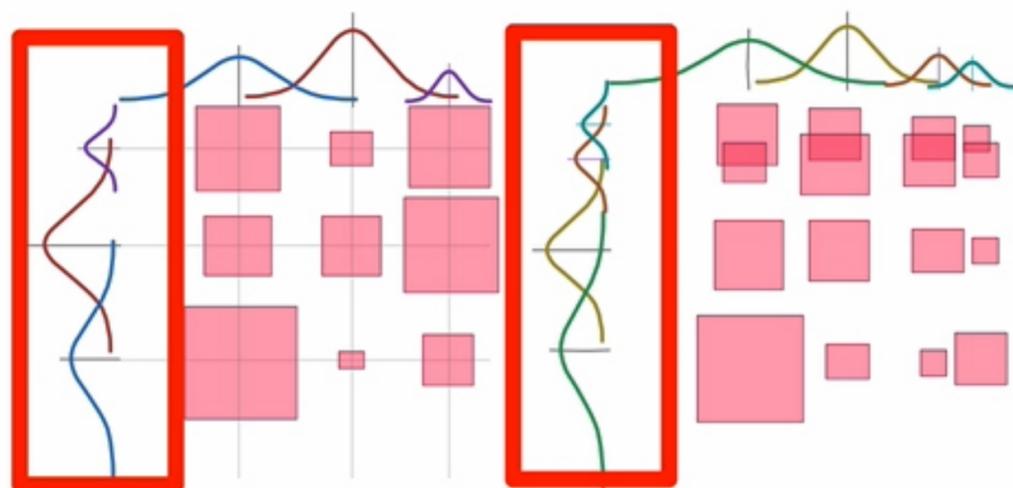
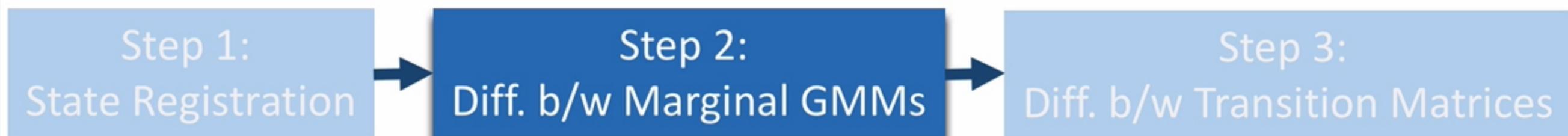


$$\min_{\mathbf{W} \in \Pi(\pi_1, \pi_2)} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w_{i,j} W^2(\phi_{1,i}, \phi_{2,j})$$

$$\Pi(\pi_1, \pi_2) \stackrel{\text{def}}{=} \left\{ \mathbf{W} \in \mathbb{R}^{M_1 \times M_2} : \sum_{i=1}^{M_1} w_{i,j} = \pi_{2,j}, j = 1, \dots, M_2; \right.$$

$$\left. \sum_{j=1}^{M_2} w_{i,j} = \pi_{1,i}, i = 1, \dots, M_1; \text{ and } w_{i,j} \geq 0, \forall i, j \right\}$$

# Minimized Aggregated Wasserstein (MAW)



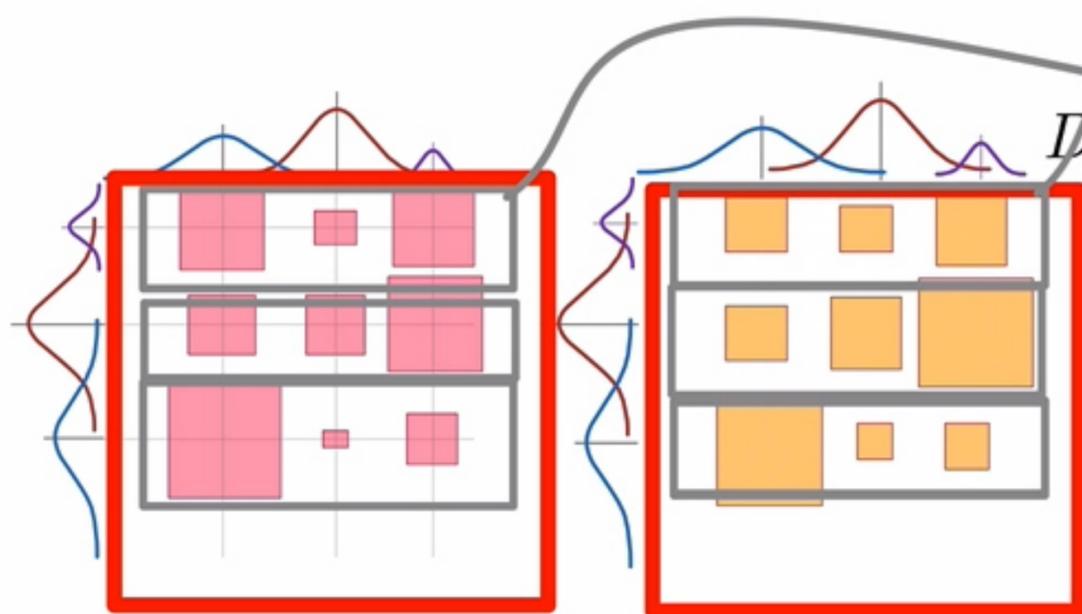
$$\widetilde{W}_0^2(\mathcal{M}_1, \mathcal{M}_2; \mathbf{W}) \stackrel{\text{def}}{=} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w_{i,j} W^2(\phi_{1,i}, \phi_{2,j})$$

# Minimized Aggregated Wasserstein (MAW)

Step 1:  
State Registration

Step 2:  
Diff. b/w Marginal GMMs

Step 3:  
Diff. b/w Transition Matrices



$$D(\mathbf{T}_1, \mathbf{T}_2 : \mathbf{W}) \stackrel{\text{def}}{=} d_T(\mathbf{T}_1, \tilde{\mathbf{T}}_2) + d_T(\mathbf{T}_2, \tilde{\mathbf{T}}_1)$$

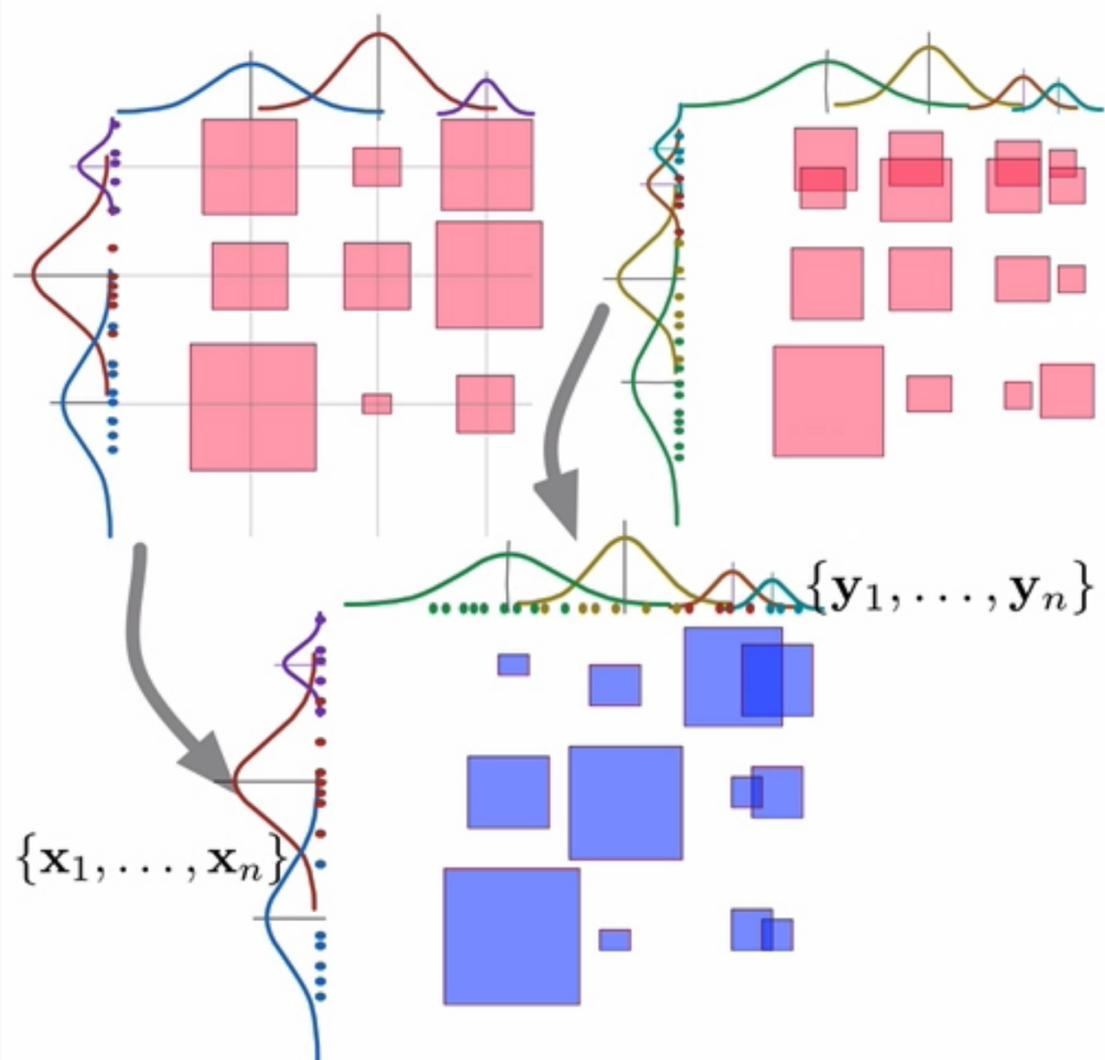
$$d_T(\mathbf{T}_1, \tilde{\mathbf{T}}_2) \stackrel{\text{def}}{=} \sum_{i=1}^{M_1} \pi_{1,i} \tilde{W}^2 \left( \mathcal{M}_1^{(i)} |_{\mathbf{T}_1(i,:)}, \mathcal{M}_1^{(i)} |_{\tilde{\mathbf{T}}_2(i,:)} \right)$$

$$d_T(\mathbf{T}_2, \tilde{\mathbf{T}}_1) \stackrel{\text{def}}{=} \sum_{i=1}^{M_2} \pi_{2,i} \tilde{W}^2 \left( \mathcal{M}_2^{(i)} |_{\mathbf{T}_2(i,:)}, \mathcal{M}_2^{(i)} |_{\tilde{\mathbf{T}}_1(i,:)} \right)$$

$$MAW(\Lambda_1, \Lambda_2) \stackrel{\text{def}}{=} (1 - \alpha) \tilde{W}_0(\mathcal{M}_1, \mathcal{M}_2; \mathbf{W}) + \alpha D(\mathbf{T}_1, \mathbf{T}_2 : \mathbf{W})$$

- Semi-metric

# Improved Aggregated Wasserstein (IAW)



$$\widetilde{\mathbf{W}}_n^* \stackrel{\text{def}}{=} [\pi(\mathbf{x}_1; \mathcal{M}_1), \dots, \pi(\mathbf{x}_n; \mathcal{M}_1)] \\ \cdot \Pi_n \cdot [\pi(\mathbf{y}_1; \mathcal{M}_2), \dots, \pi(\mathbf{y}_n; \mathcal{M}_2)]^T$$

Optimal Coupling of sampled points

Speed up of computation:

KL-regularized Optimization [Cuturi *et al.* NIPS13],

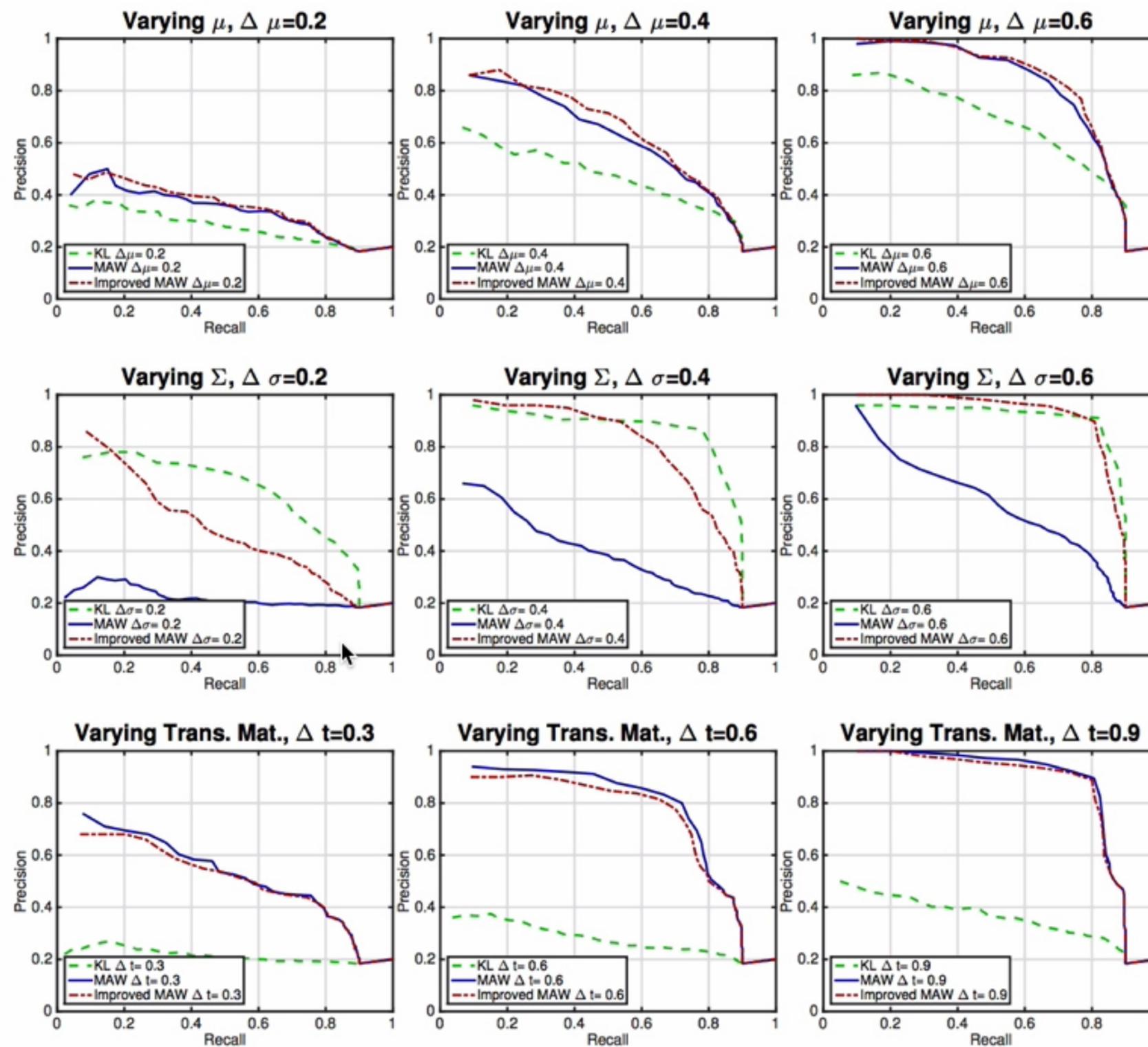
Bregman ADMM (on GPU) [Huang *et al.*, NIPS14]

# Experiments:

Synthetic data experiment:

Real data experiment:

Sequence  
Nearest neighbor  
Retrieval:



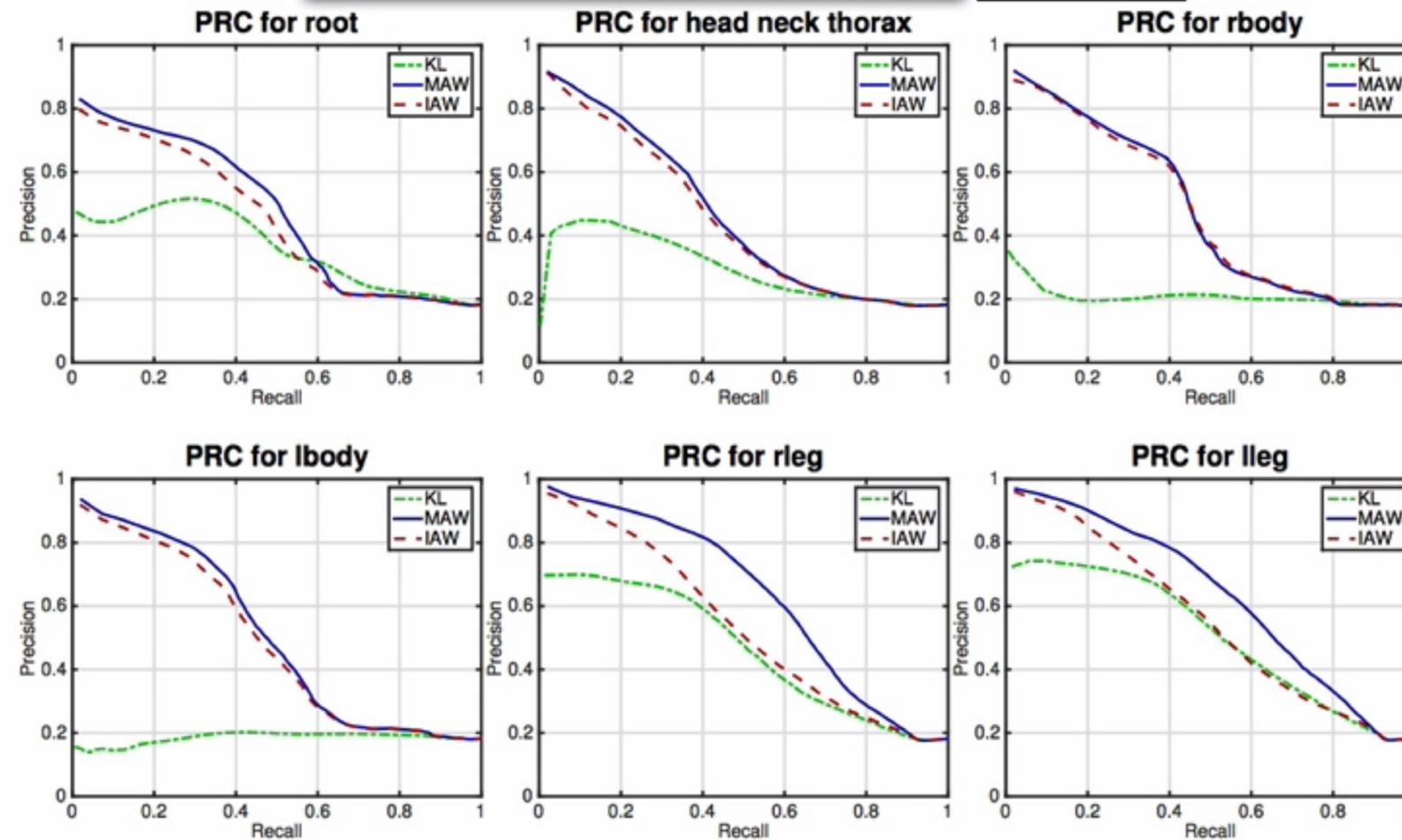
# Experiments:

Synthetic data experiment:

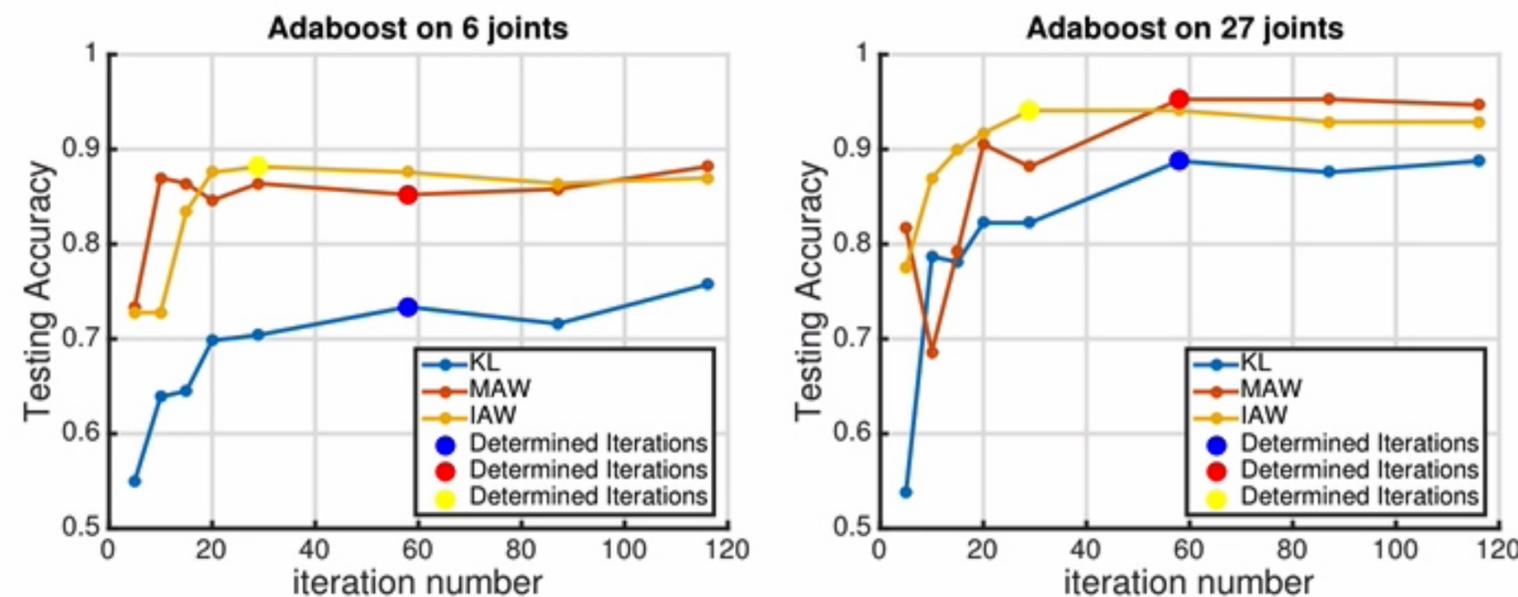
Real data experiment:



Sequence  
Nearest neighbor  
Retrieval:



Sequence  
Adaboost  
Classification:



Thank you!

For more details, please refer to our paper.

And welcome to come to our poster spot.