Carnegie Mellon
School of Computer Science

Time Series Clustering: Complex is Simpler!

Lei Li

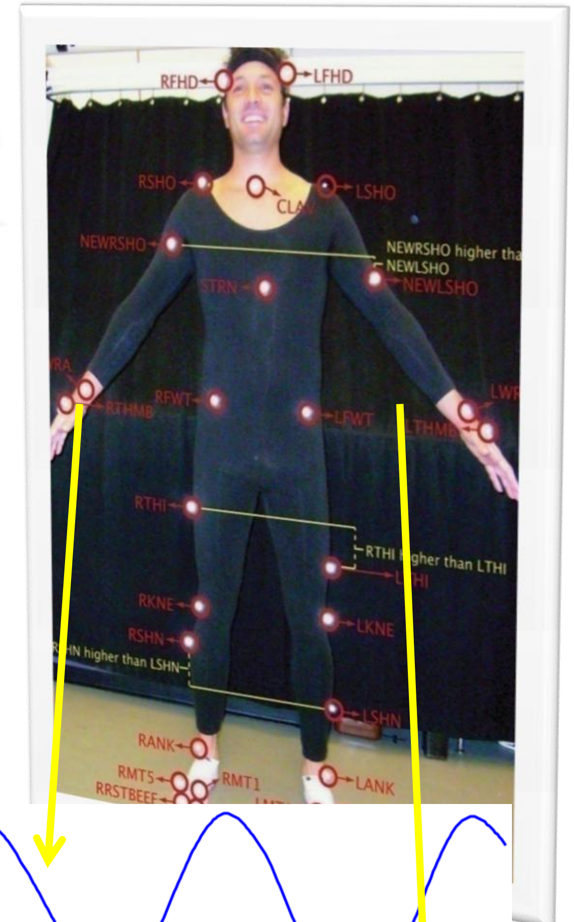
B. Aditya Prakash

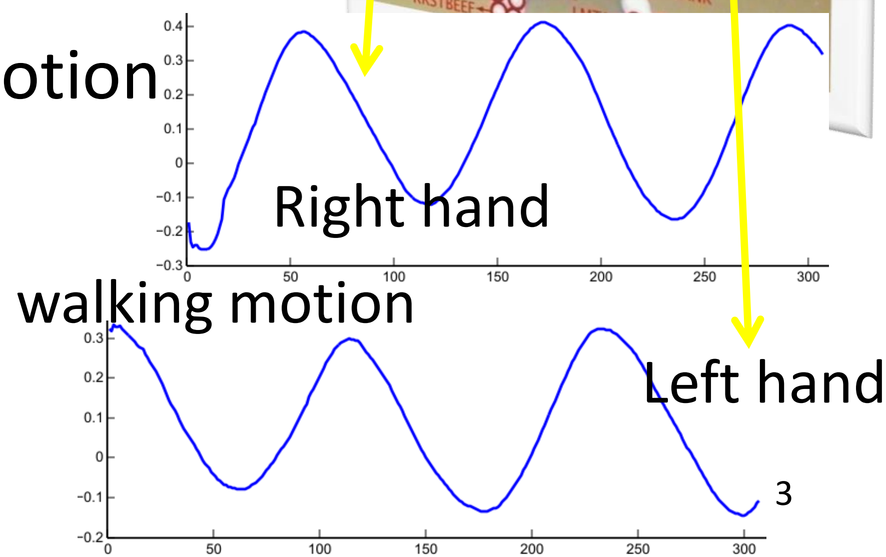
leili,badityap@cs.cmu.edu

Why time series clustering?

Motion Capture

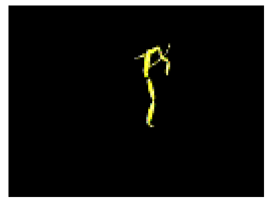
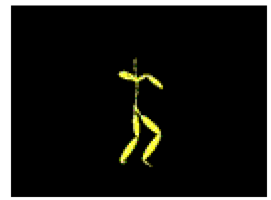
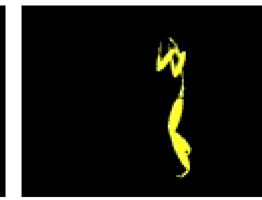
- Application of motion capture
  - Game ($57B)
  - Movie industry
- Goal:
  - Understand human motion
  - Generate new natural motion [Li et al, 2008a]
- Sub-goal:
  - automatic labeling





Answering similarity queries

[Li et al, 2010]

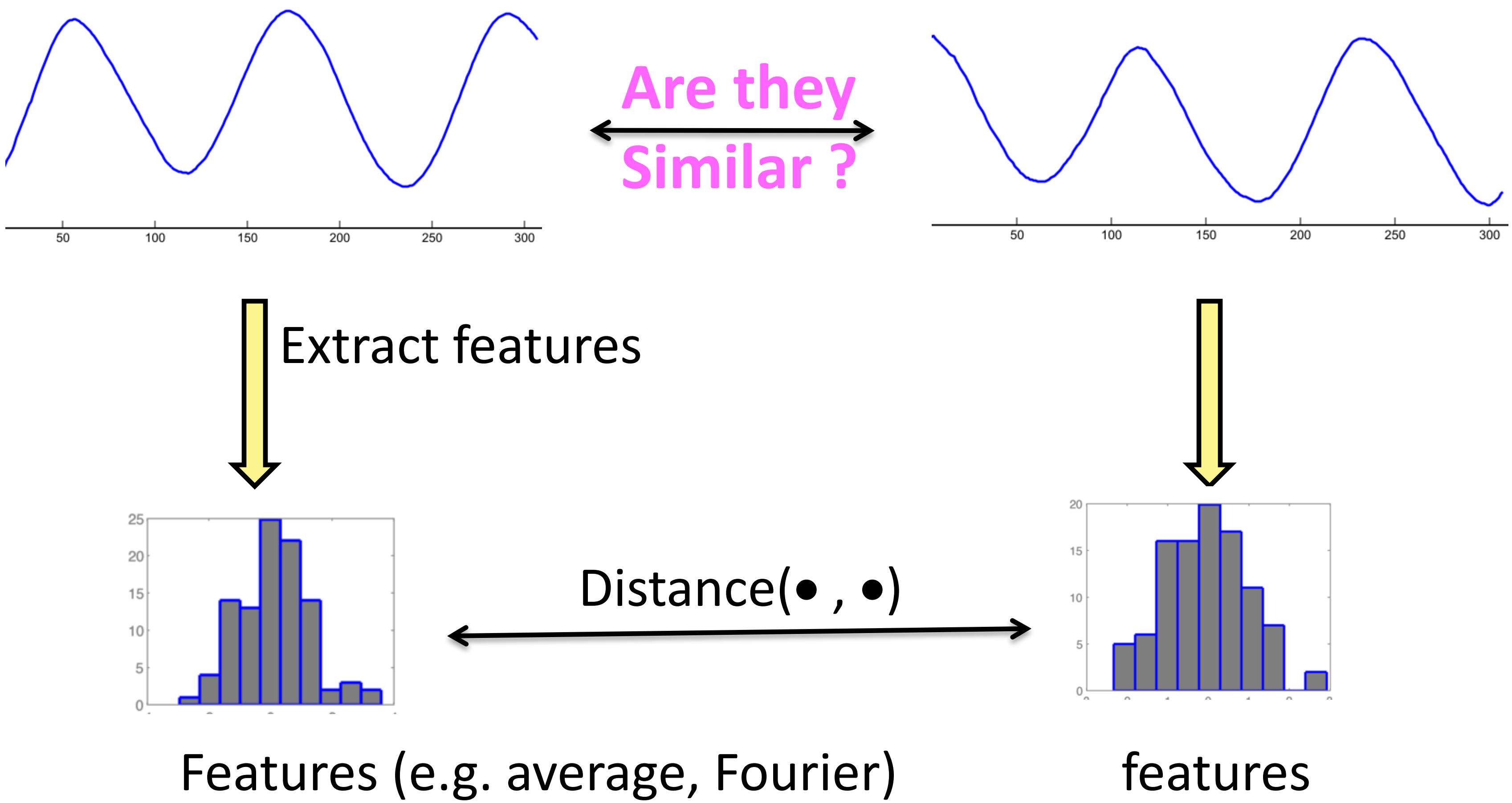
SELECT * **FROM**    

WHERE time_seq.

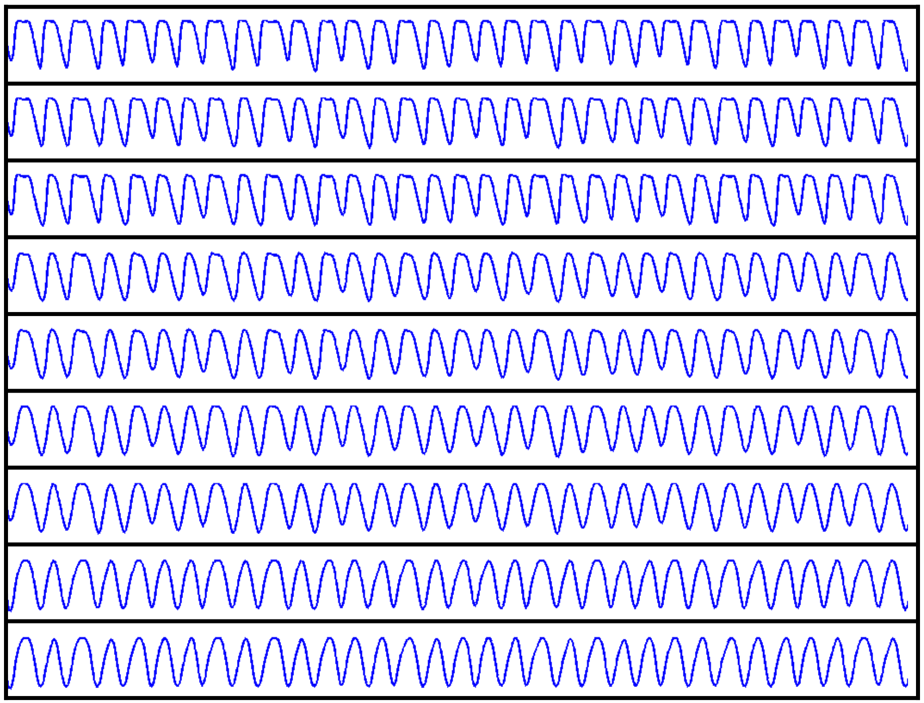
LIKE 

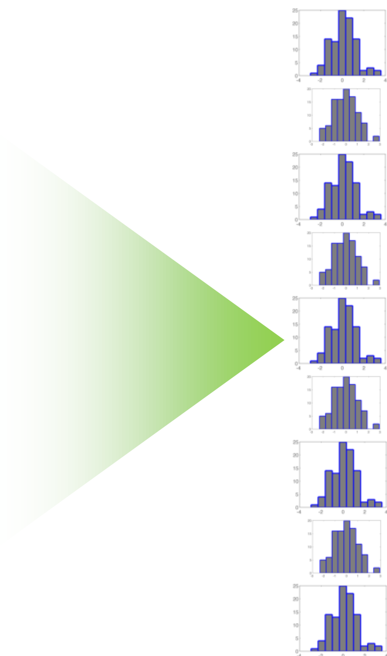
Central Problem

- Estimate “Similarity” among time sequences



What are good features?





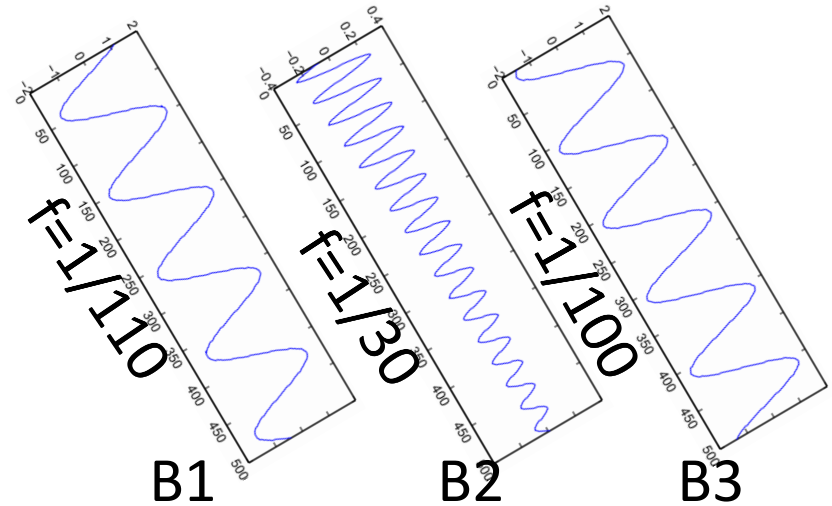
Requirements of good features:

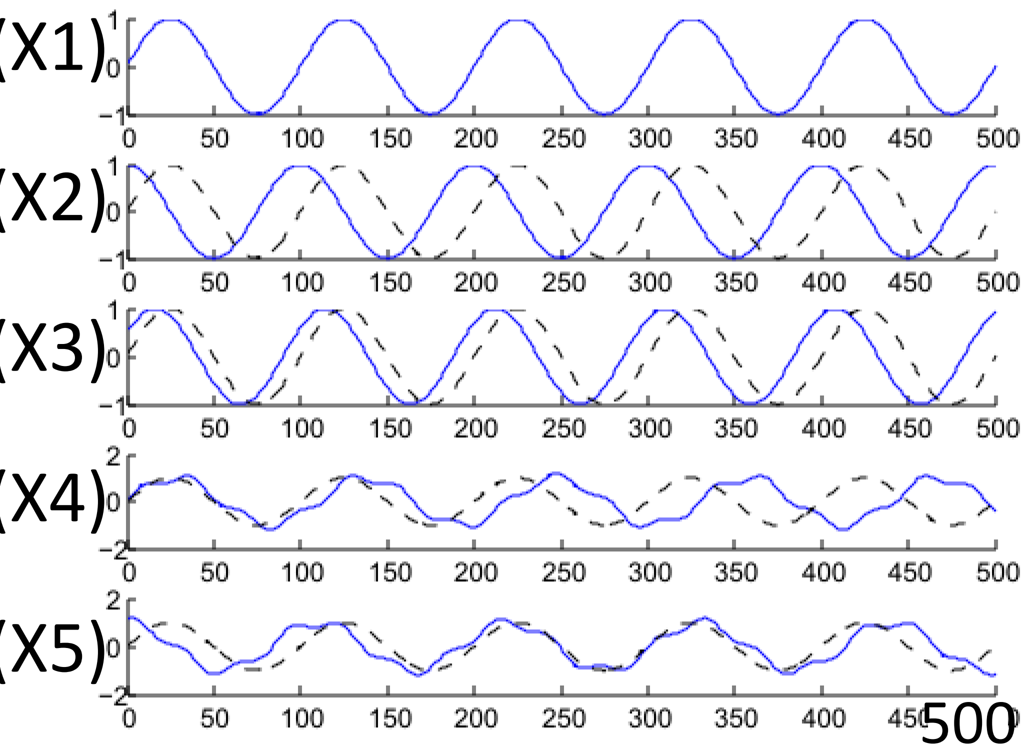
0. ***Agree*** with human intuition
1. Time lag
2. Frequency proximity
3. more (next)

e.g. Mocap sequences
Chlorine measurements in water
temperatures in machine room

Basic idea

learning basis/harmonics

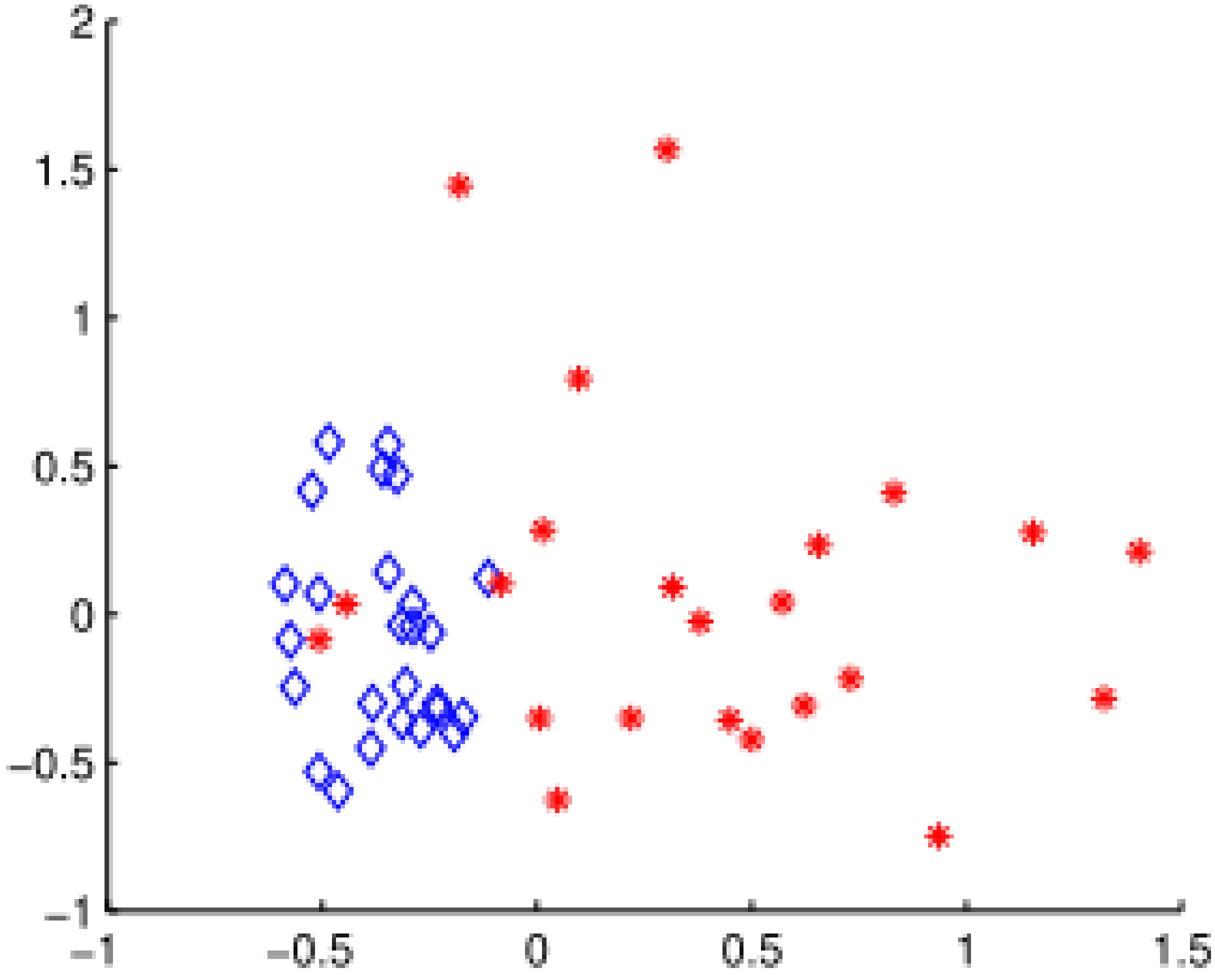




		B1		B2		B3
(X1)	=	0	+	1.0	+	0
(X2)	=	0	+	1.0	+	0
(X3)	=	0	+	0.9	+	0
(X4)	=	1.0	+	0	+	1.0
(X5)	=	1.0	+	0	+	1.0

Mixing weights

Preview of CLDS Result



• running

◊ walking

Outline

- Motivation
- Background: Complex Normal Distribution
- Complex Linear Dynamical Systems
- Clustering with CLDS
- Experiments
- Results

Complex Normal Distribution

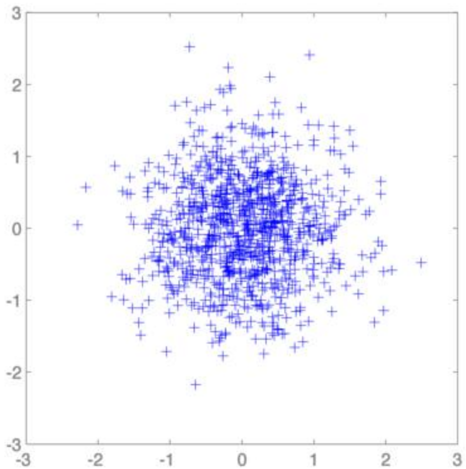
- Example: $x = a + ib$

  standard complex normal distribution

$$x \sim CN(0,1) \quad \longleftrightarrow \quad p(x) = \frac{1}{\pi} e^{-|x|^2}$$

$$\binom{a}{b} \sim N\left(\binom{0}{0}, \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \longleftrightarrow \quad p(a,b)$$

$$= (2\pi)^{-1} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\binom{a}{b}-\mu\right)' \Sigma^{-1}\left(\binom{a}{b}-\mu\right)}$$



Complex Normal Distribution

- x is said to follow the complex normal distribution, if its p.d.f

$\boldsymbol{x} \sim \mathcal{CN}(\mu, H),\ if\ its\ p.d.f\ is$

$$p(\boldsymbol{x}) = \pi^{-m}|H|^{-1}\exp(-(\boldsymbol{x}-\mu)^{*}H^{-1}(\boldsymbol{x}-\mu))$$

H is hermitian matrix, $(\cdot)^{*}$ is conjugate transpose

[Goodman, 1963]

Compare to Normal Distribution

Complex Normal Distribution

$\boldsymbol{x} \sim \mathcal{CN}(\mu, H),\ if\ its\ p.d.f\ is$

$$p(\boldsymbol{x}) = \pi^{-m}|H|^{-1}\exp(-(\boldsymbol{x}-\mu)^{*}H^{-1}(\boldsymbol{x}-\mu))$$

H is hermitian matrix, $(\cdot)^{*}$ is conjugate transpose

Normal Distribution

$\boldsymbol{x} \sim \mathcal{N}(\mu, \Sigma),\ if\ its\ p.d.f\ is$

$$p(\boldsymbol{x}) = (2\pi)^{-m/2}|\Sigma|^{1/2}\exp(-\frac{(\boldsymbol{x}-\mu)^{T}\Sigma^{-1}(\boldsymbol{x}-\mu)}{2})$$

H is hermitian matrix, $(\cdot)^{*}$ is conjugate transpose

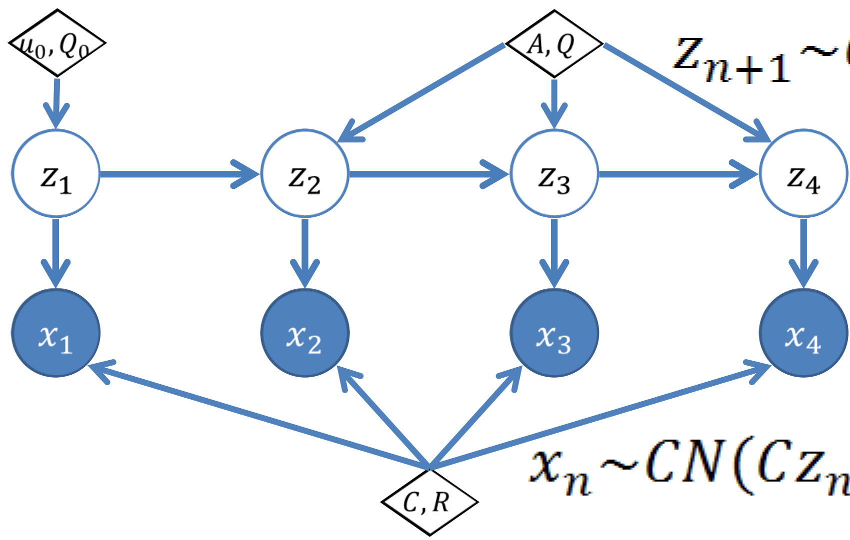
Outline

- Motivation
- Background: Complex Normal Distribution
- Complex Linear Dynamical Systems
- Clustering with CLDS
- Experiments
- Conclusion

(Complex) Linear Dynamical Systems

$$z_1 \sim CN(\mu_0, Q_0)$$

$$z_{n+1} \sim CN(Az_n, Q)$$

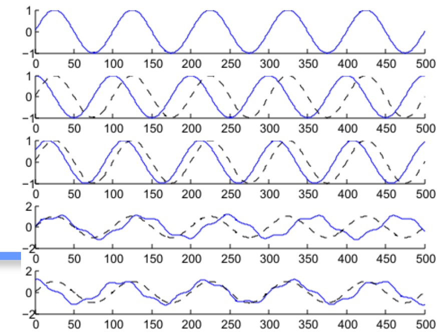


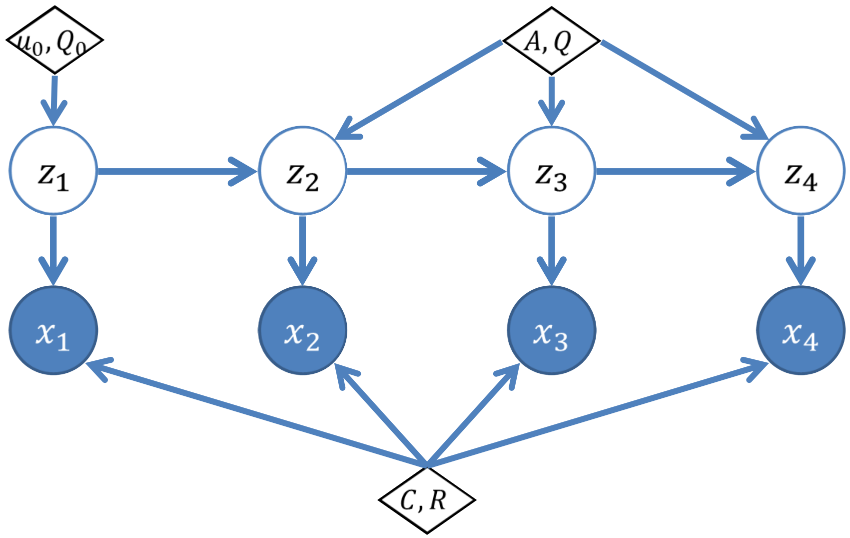
$$x_n \sim CN(Cz_n, R)$$

$$\mu_0 = \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix} \qquad A = \begin{pmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{pmatrix} \qquad C = \begin{pmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{pmatrix}$$

$$Q_0 = \begin{pmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{pmatrix} \qquad Q = \begin{pmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{pmatrix} \qquad R = \begin{pmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{pmatrix}$$

Example





$$z_1 \sim CN(\mu_0, Q_0)$$
$$z_{n+1} \sim CN(Az_n, Q)$$
$$x_n \sim CN(Cz_n, R)$$

A: transition matrix

C: output matrix

$$A = \begin{pmatrix} 0.9984 + 0.0571i & 0 & 0 \\ 0 & 0.9980 + 0.0628i & 0 \\ 0 & 0 & 0.9781 + 0.2079i \end{pmatrix}$$

$$C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & i & 0 \\ 0 & 0.866 + 0.5i & 0 \\ 1 & 0 & 1 \\ i & 0 & 0.707 + 0.707i \end{pmatrix}$$

Parameter Learning

min
$$\mathcal{L}(\theta) = \mathbb{E}_{\boldsymbol{Z}|\boldsymbol{X}}[-\log P(\boldsymbol{X},\boldsymbol{Z}|\theta)]$$
$$= \log|\boldsymbol{Q}_0| + \mathbb{E}[(\boldsymbol{z}_1 - \boldsymbol{\mu}_0)^* \boldsymbol{Q}_0^{-1}(\boldsymbol{z}_1 - \boldsymbol{\mu}_0)]$$
$$+ \mathbb{E}[\sum_{n=1}^{N-1}(\boldsymbol{z}_{n+1} - \boldsymbol{A}\cdot\boldsymbol{z}_n)^* \cdot Q^{-1} \cdot (\boldsymbol{z}_{n+1} - \boldsymbol{A}\cdot\boldsymbol{z}_n)] + (N-1)\log|Q|$$
$$+ \mathbb{E}[\sum_{n=1}^{N}(\boldsymbol{x}_n - \boldsymbol{C}\cdot\boldsymbol{z}_n)^* \cdot R^{-1} \cdot (\boldsymbol{x}_n - \boldsymbol{C}\cdot\boldsymbol{z}_n)] + N\log|R|$$

EM algorithm (complex-Fit)

- E-step: compute posterior $P(z_n|x_1, \ldots, x_N)$ and $P(z_n, z_{n+1}|x_1, \ldots, x_N)$
- M-step: update the parameters to optimize $L(\theta)$

Optimizing real-valued functions of complex variables

- With real variables:
  - $\frac{df}{dx} = 0$
  - Gradient descent: $x \leftarrow x - \alpha f'$
- With complex variables:
  - $\frac{\partial f}{\partial x} = 0$ AND $\frac{\partial f}{\partial \bar{x}} = 0$
- EM algorithm (complex-Fit)

$$\frac{\partial}{\partial \boldsymbol{\mu}_0}\mathcal{L} = 0 \quad \frac{\partial}{\partial \overline{\boldsymbol{\mu}_0}}\mathcal{L} = 0 \quad \frac{\partial}{\partial \boldsymbol{Q}_0}\mathcal{L} = 0 \quad \frac{\partial}{\partial \overline{\boldsymbol{Q}_0}}\mathcal{L} = 0$$

$$\frac{\partial}{\partial \mathbf{A}}\mathcal{L}, \frac{\partial}{\partial \overline{\mathbf{A}}}\mathcal{L}, \frac{\partial}{\partial \mathbf{Q}}\mathcal{L}, \frac{\partial}{\partial \overline{\mathbf{Q}}}\mathcal{L}, \frac{\partial}{\partial \mathbf{C}}\mathcal{L}, \frac{\partial}{\partial \overline{\mathbf{C}}}\mathcal{L}, \frac{\partial}{\partial \mathbf{R}}\mathcal{L}, \frac{\partial}{\partial \overline{\mathbf{R}}}\mathcal{L} = 0$$

CLDS versus LDS

CLDS

$$\min \ \mathcal{L}(\theta) = \mathbb{E}_{\boldsymbol{Z}|\boldsymbol{X}}[-\log P(\boldsymbol{X}, \boldsymbol{Z}|\theta)]$$

$$= \log|\boldsymbol{Q}_0| + \mathbb{E}[(\boldsymbol{z}_1 - \boldsymbol{\mu}_0)^* \boldsymbol{Q}_0^{-1} (\boldsymbol{z}_1 - \boldsymbol{\mu}_0)]$$

$$+ \mathbb{E}[\sum_{n=1}^{N-1} (\boldsymbol{z}_{n+1} - \boldsymbol{A} \cdot \boldsymbol{z}_n)^* \cdot Q^{-1} \cdot (\boldsymbol{z}_{n+1} - \boldsymbol{A} \cdot \boldsymbol{z}_n)] + (N-1) \log|Q|$$

$$+ \mathbb{E}[\sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{C} \cdot \boldsymbol{z}_n)^* \cdot R^{-1} \cdot (\boldsymbol{x}_n - \boldsymbol{C} \cdot \boldsymbol{z}_n)] + N \log|R|$$

LDS

$$\min \ \mathcal{L}(\theta) = \mathbb{E}_{\boldsymbol{Z}|\boldsymbol{X}}[-\log P(\boldsymbol{X}, \boldsymbol{Z}|\theta)]$$

$$= \log|\boldsymbol{Q}_0| + \mathbb{E}[(\boldsymbol{z}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{Q}_0^{-1} (\boldsymbol{z}_1 - \boldsymbol{\mu}_0)]$$

$$+ \mathbb{E}[\sum_{n=1}^{N-1} (\boldsymbol{z} - \boldsymbol{A} \cdot \boldsymbol{z}_n)^T \cdot Q^{-1} \cdot (\boldsymbol{z}_{n+1} - \boldsymbol{A} \cdot \boldsymbol{z}_n)] + (N-1) \log|Q|$$

$$+ \mathbb{E}[\sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{C} \cdot \boldsymbol{z}_n)^T \cdot R^{-1} \cdot (\boldsymbol{x}_n - \boldsymbol{C} \cdot \boldsymbol{z}_n)] + N \log|R|$$

Outline

- Motivation
- Background: Complex Normal Distribution
- Complex Linear Dynamical Systems
- Clustering with CLDS
- Experiments
- Conclusion

CLDS for Clustering

Enforcing **A** to be diagonal, $A = \text{diag}(a)$

for learning **a** and **Q**:

$$\boldsymbol{a} = (Q^{-1} \circ (\sum_{n=1}^{N-1} \mathbb{E}[\boldsymbol{z}_n \cdot \boldsymbol{z}_n^*])^T)^{-1} \cdot (Q^{-1} \circ (\sum_{n=1}^{N-1} \mathbb{E}[\boldsymbol{z}_{n+1} \cdot \boldsymbol{z}_n^*])^T) \cdot \mathbf{1}$$

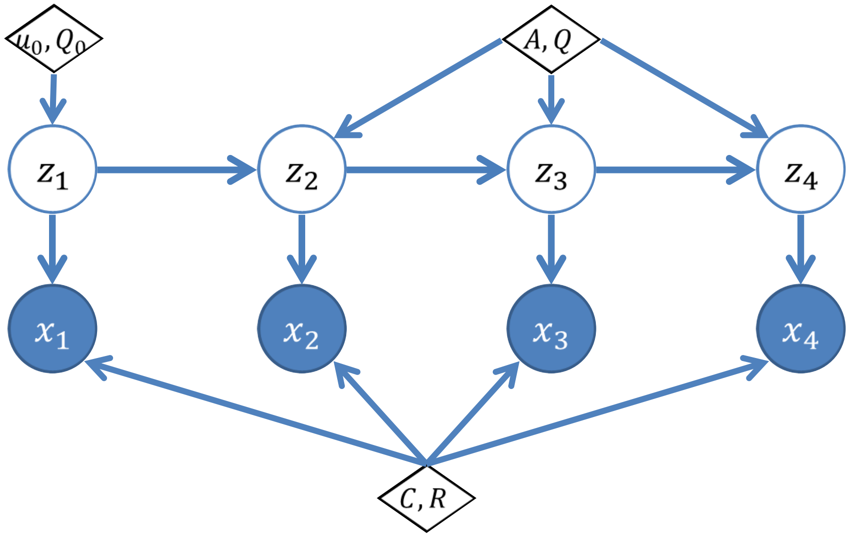
$$Q = \frac{1}{N-1} \sum_{n=1}^{N-1} \Big(\mathbb{E}[\boldsymbol{z}_{n+1} \cdot \boldsymbol{z}_{n+1}^*] - \mathbb{E}[\boldsymbol{z}_{n+1} \cdot (\boldsymbol{a} \circ \boldsymbol{z}_n)^*] - \mathbb{E}[(\boldsymbol{a} \circ \boldsymbol{z}_n) \cdot \boldsymbol{z}_{n+1}^*] + \mathbb{E}[(\boldsymbol{a} \circ \boldsymbol{z}_n) \cdot (\boldsymbol{a} \circ \boldsymbol{z}_n)^*] \Big)$$

Rationale:

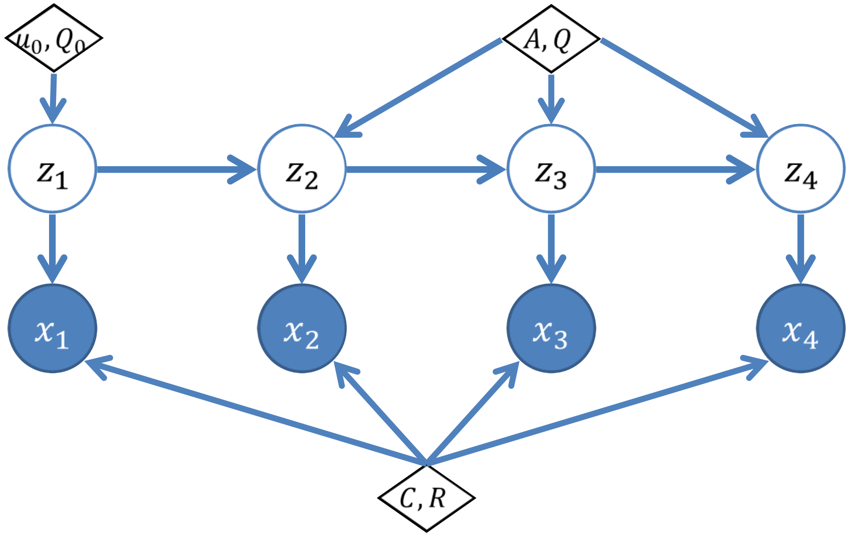
- Faster
- More robust
- Better clustering

Features

- z's will be basis
- C will contain features
- To eliminate time shift, take magnitude of C



Example



$$z_1 \sim CN(\mu_0, Q_0)$$
$$z_{n+1} \sim CN(Az_n, Q)$$
$$x_n \sim CN(Cz_n, R)$$

A: transition matrix

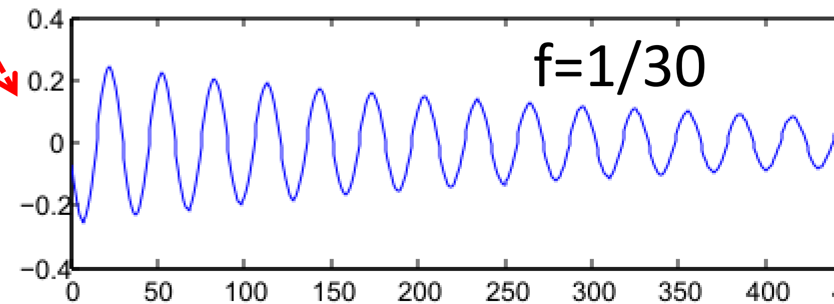
C: output matrix

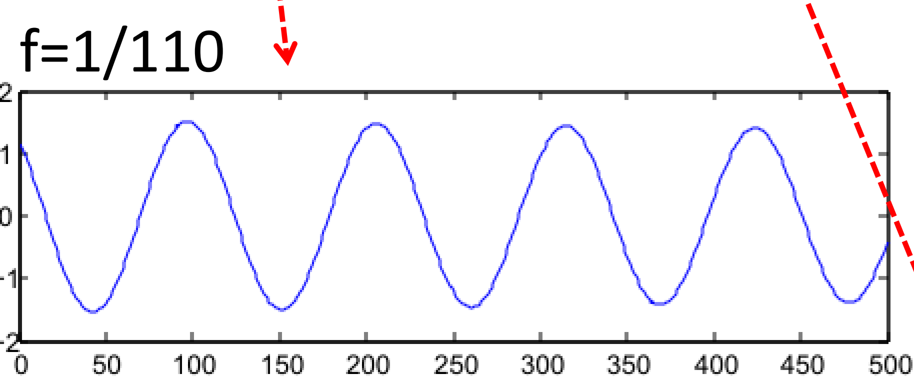
$$A = \begin{pmatrix} 0.9984 + 0.0571i & 0 & 0 \\ 0 & 0.9980 + 0.0628i & 0 \\ 0 & 0 & 0.9781 + 0.2079i \end{pmatrix}$$

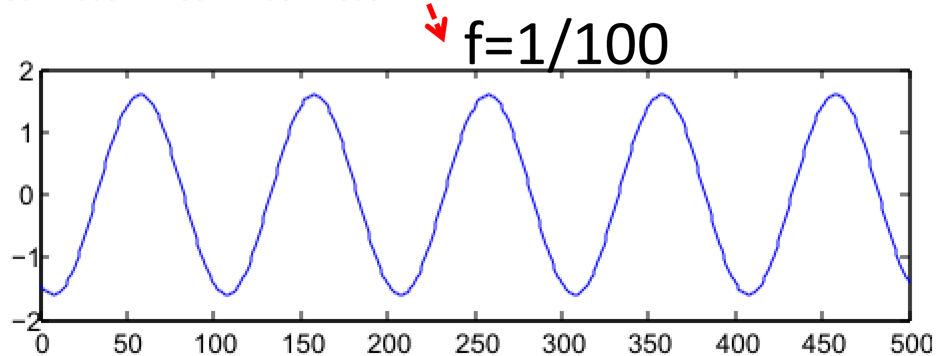
$$C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & i & 0 \\ 0 & 0.866 + 0.5i & 0 \\ 1 & 0 & 1 \\ i & 0 & 0.707 + 0.707i \end{pmatrix}$$

Simple interpretation for "Complex" solution

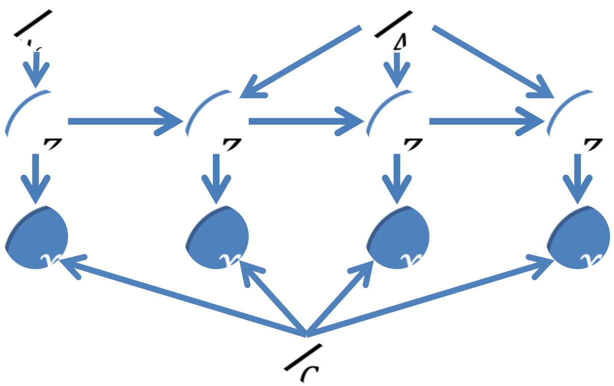
$$A = \begin{pmatrix} 0.9984 + 0.0571i & 0 & 0 \\ 0 & 0.9980 + 0.0628i & 0 \\ 0 & 0 & 0.9781 + 0.2079i \end{pmatrix}$$







Simple interpretation for “Complex” solution



$$C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & i & 0 \\ 0 & 0.866 + 0.5i & 0 \\ 1 & 0 & 1 \\ i & 0 & 0.707 + 0.707i \end{pmatrix}$$

Take magnitude

Feature Matrix F

0	1	0
0	1	0
0	1	0
1	0	1
1	0	1

Outline

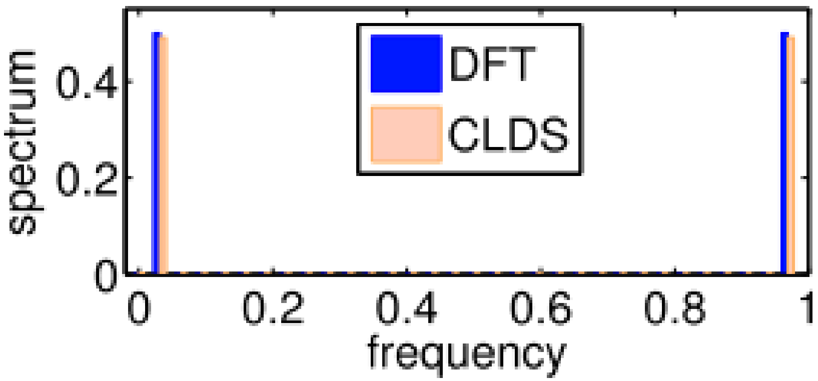
- Motivation
- Background: Complex Normal Distribution
- Complex Linear Dynamical Systems
- Clustering with CLDS
- Experiments
- Conclusion

DFT as a special case

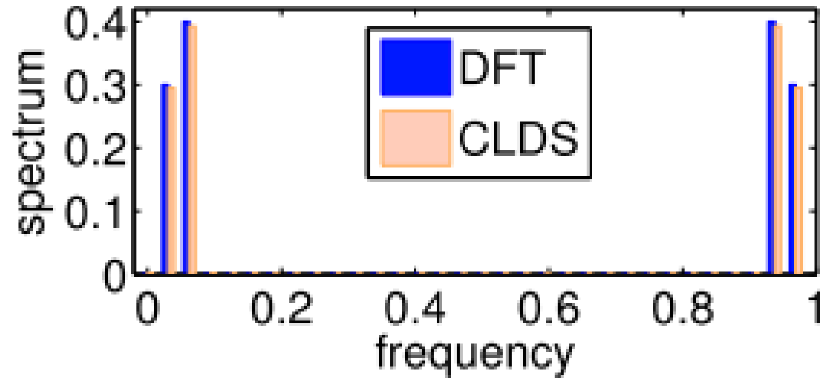
For single signal,

If $\boldsymbol{A} = \text{diag}(\exp(\frac{2\pi i}{N} k)), k = 1, \ldots, N$

C will be Fourier specturm



(a) $x = \sin(\frac{2\pi t}{32})$



(b) $x = 0.6 \sin(\frac{2\pi t}{32}) + 0.8 \sin(\frac{2\pi t}{16})$

Results

- Datasets:
  - MOCAPPOS:
    - 49 motion sequences
    - marker positions
    - running v.s. walking
  - MOCAPANG:
    - 33 sequences
    - joint angles
- Metric: conditional entropy of the confusion matrix M

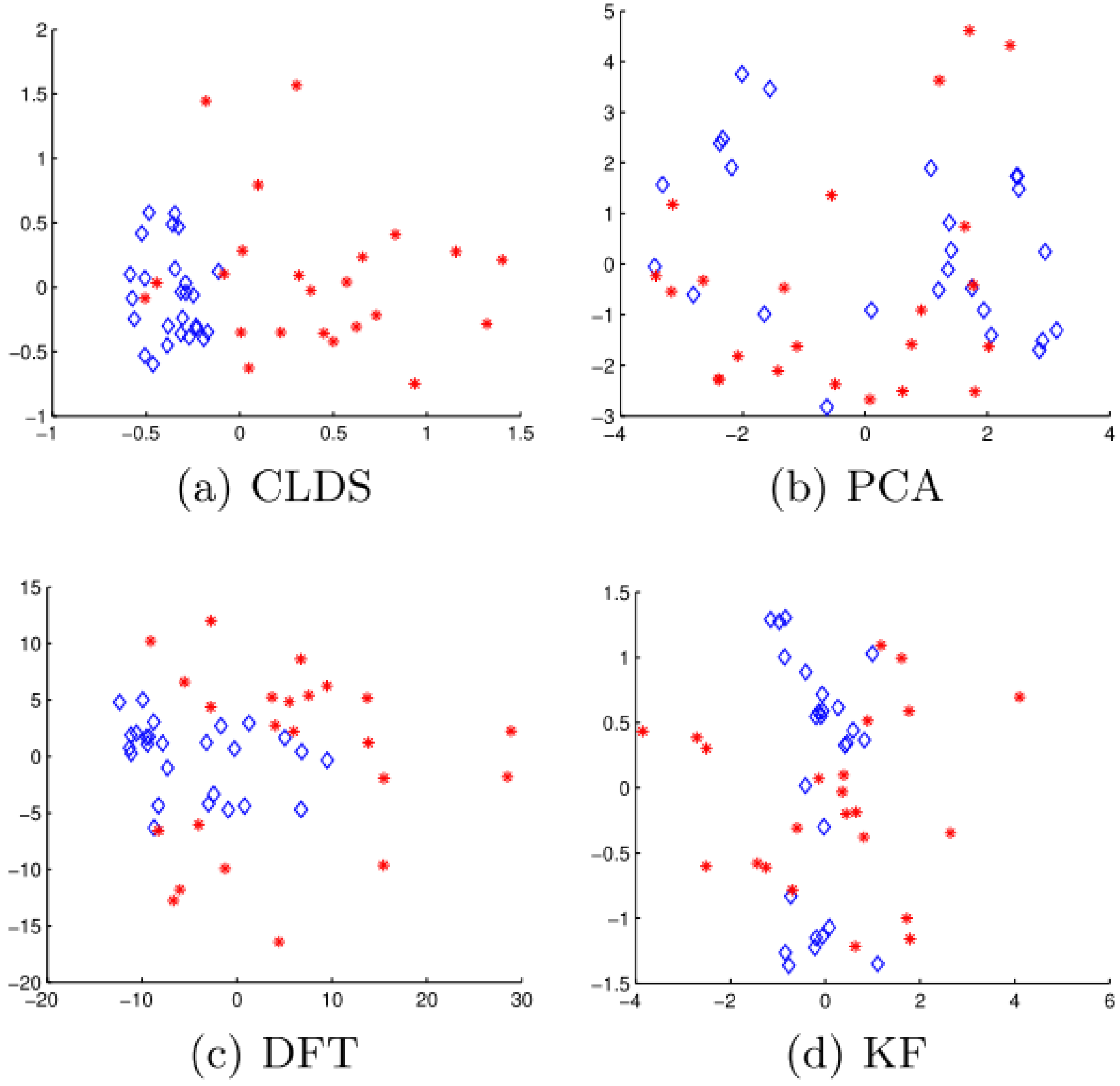
$$S(M) = \sum_{i,j} \frac{M_{i,j}}{\sum_{k,l} M_{k,l}} \log \frac{\sum_k M_{i,k}}{M_{i,j}}$$

Results

Conditional Entropy

methods	MOCAPPOS S	MOCAPANG S	
CLDS	**0.3786**	**0.1015**	
PCA	0.6818	0.3635	[Bishop 2006]
DFT	0.6143	0.2538	
DTW	0.5707	0.4229	[Gunopulos 2001]
KF	0.6749	0.5239	[Buzan 2004]

Visualization of CLDS Features

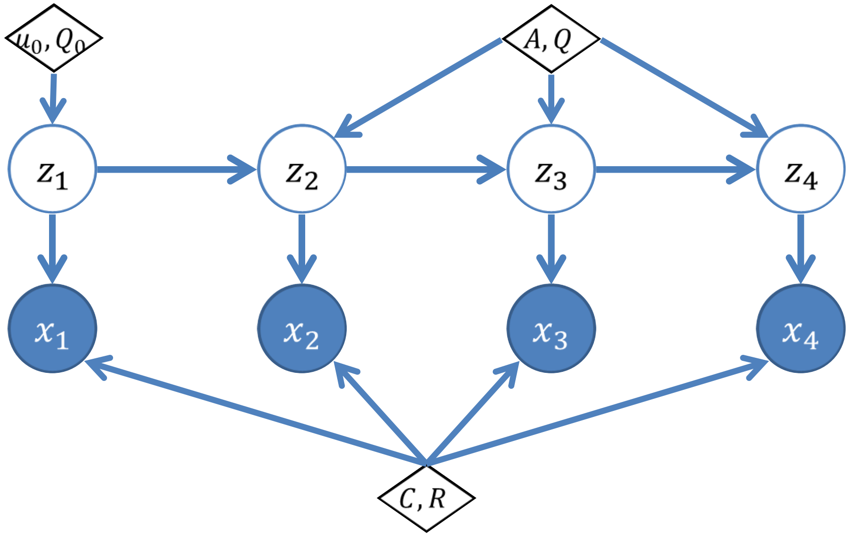


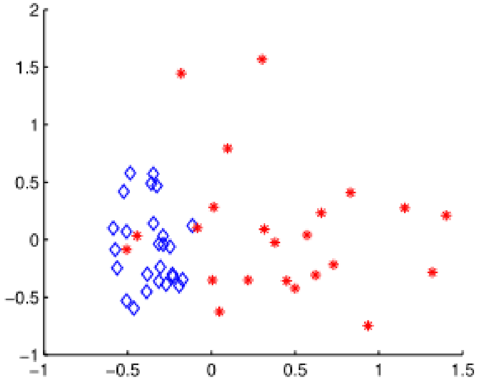
Conclusion

- Proposed Complex Linear Dynamical Systems
  - *Complex Normal* Distributions for lag and harmonics
  - *Diagonal* transition matrix for time series clustering
- Complex-Fit for learning parameters
- Advantages:
  - Faster
  - More robust
  - Better clustering

Thanks!

- www.cs.cmu.edu/~leili
- Lei Li, leili@cs.cmu.edu
- B. Aditya Prakash, badityap@cs.cmu.edu





(a) CLDS