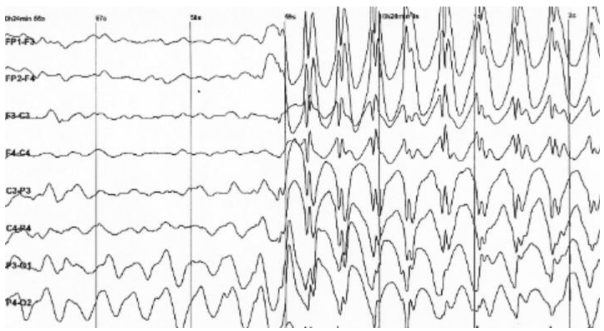
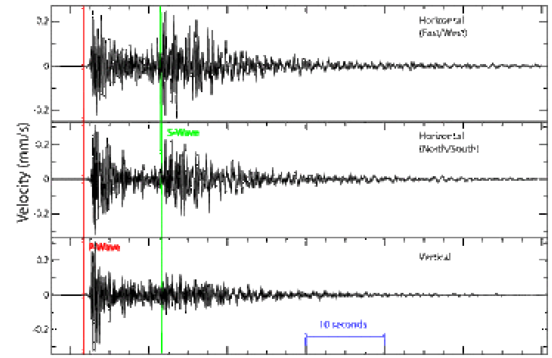
Mode Clustering for Markov Jump Systems

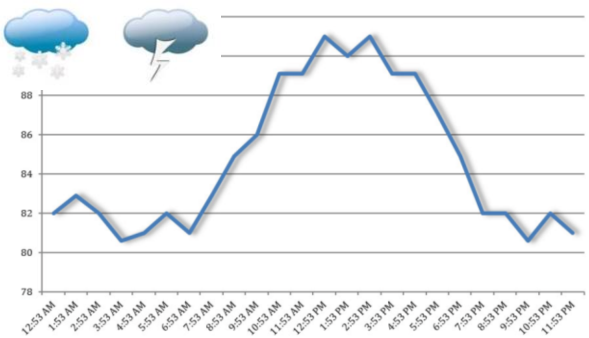
Zhe Du, Laura Balzano, Necmiye Ozay

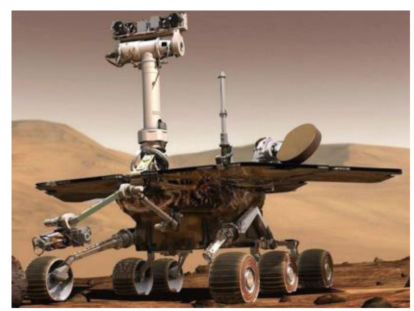
ECE, University of Michigan

Background

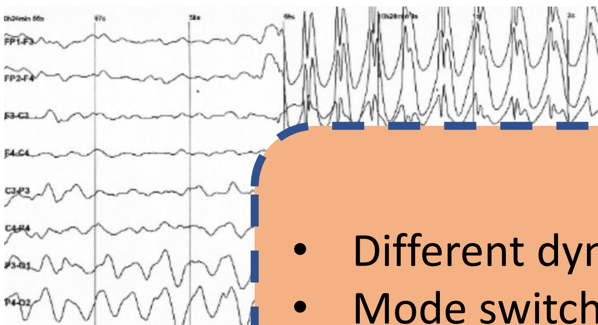

FP1-F3
FP2-F4
F3-C3
F4-C4
C3-P3
C4-P4
P3-O1
P4-O2

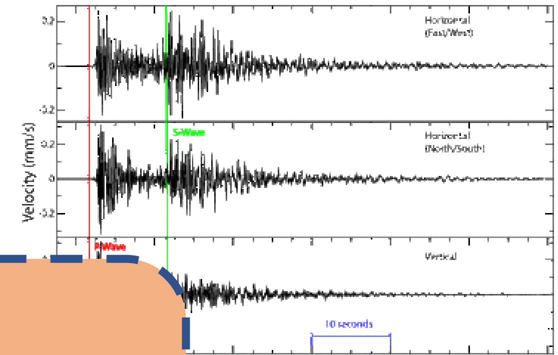

Velocity (mm/s)
Horizontal (East/West)
S-Wave
Horizontal (North/South)
P-Wave
Vertical
10 seconds


88
86
84
82
80
78
12:53 AM
1:53 AM
2:53 AM
3:53 AM
4:53 AM
5:53 AM
6:53 AM
7:53 AM
8:53 AM
9:53 AM
10:53 AM
11:53 AM
12:53 PM
1:53 PM
2:53 PM
3:53 PM
4:53 PM
5:53 PM
6:53 PM
7:53 PM
8:53 PM
9:53 PM
10:53 PM
11:53 PM

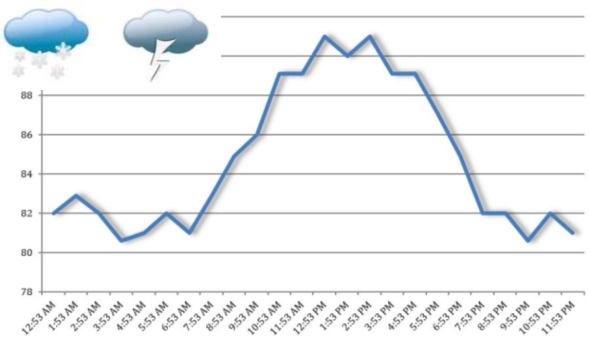


Markov Jump System



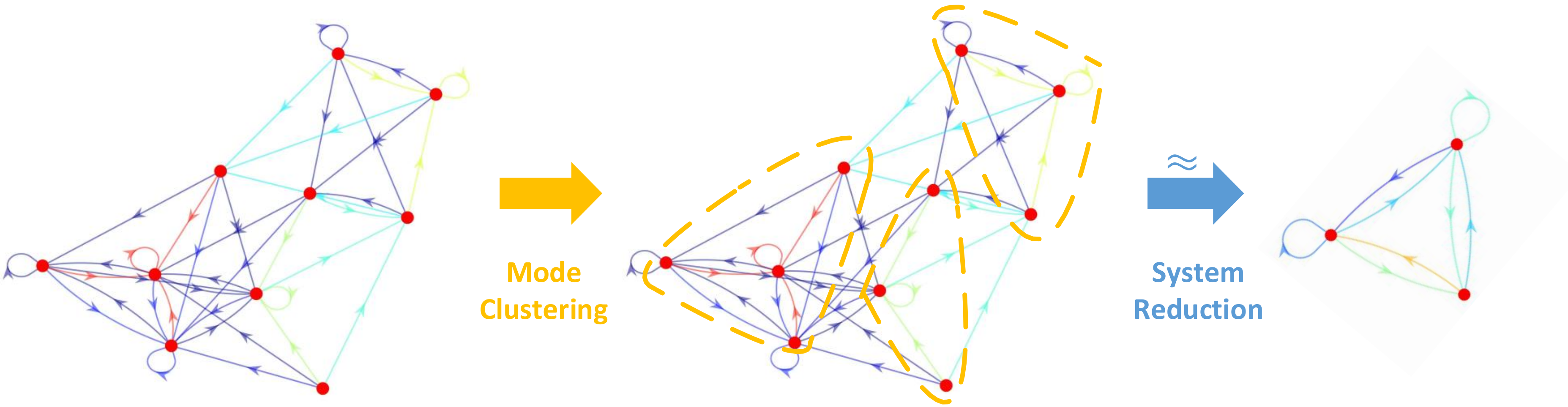


- Different dynamics under different modes
- Mode switching model-> Markov chain
- But, the state space of Markov chain can be *huge*



ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY OF MICHIGAN

Motivation



n modes $\to$ r clusters

r clusters $\to$ r modes

Benefits:

- Reveal mode relations
- Power method: $O(n^2) \to O(rn)$
  - Compute stationary distribution
  - Predict mode evolution in model predictive control (MPC)

- Reduce complexity in control planning
  - Linear quadratic regulator (LQR): $O(n^2) \to O(r^2)$

Markov Jump System

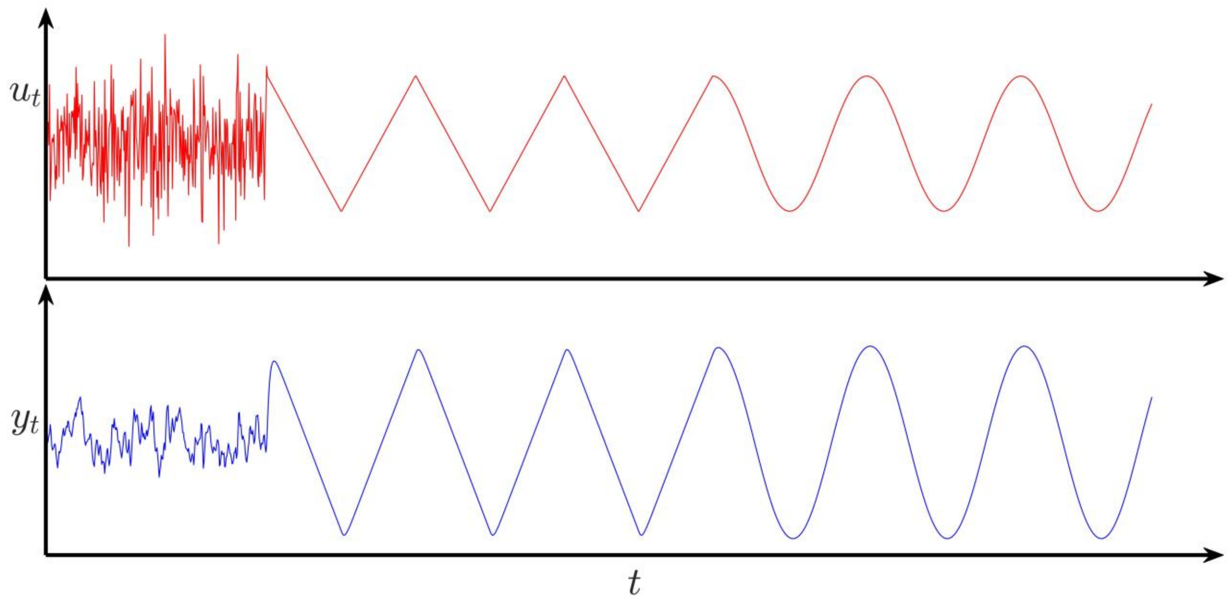
dynamics

disturbance

$$y_t = \sum_{i=1}^{n_a} a_i y_{t-i} + \sum_{j=1}^{n_c} c_j u_{t-j} + n_t$$

output

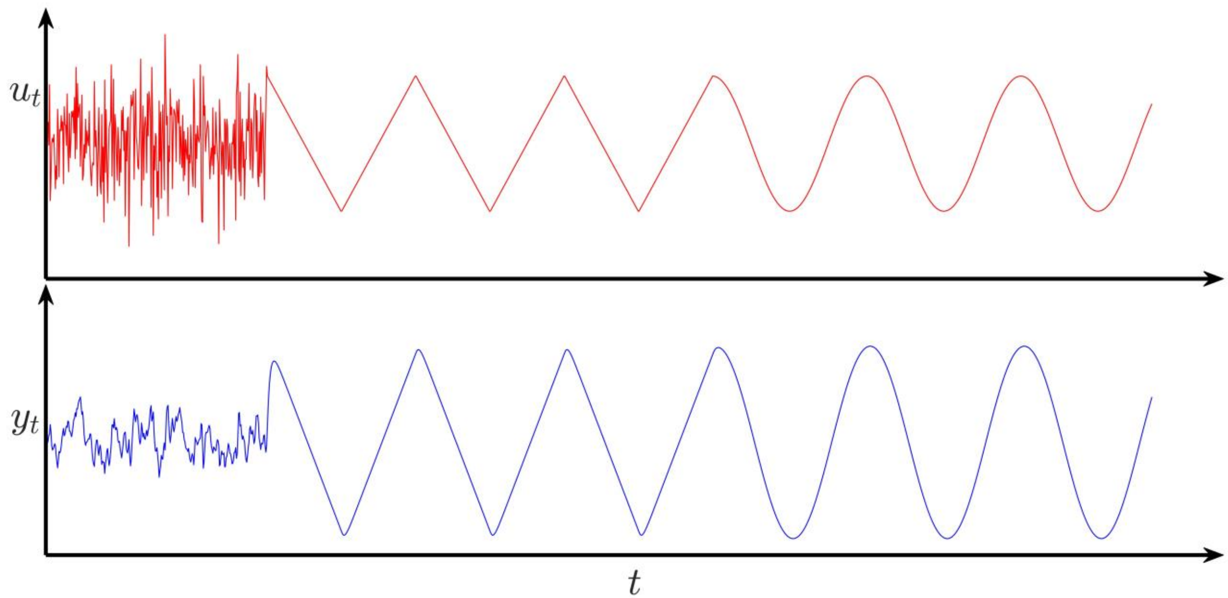
input



Markov Jump System

$$y_t = \mathbf{w}^\top \boldsymbol{\phi}_t + n_t$$

$$\boldsymbol{\phi}_t = \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-n_a} \\ u_{t-1} \\ \vdots \\ u_{t-n_c} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} a_1 \\ \vdots \\ a_{n_a} \\ c_1 \\ \vdots \\ c_{n_c} \end{bmatrix}$$

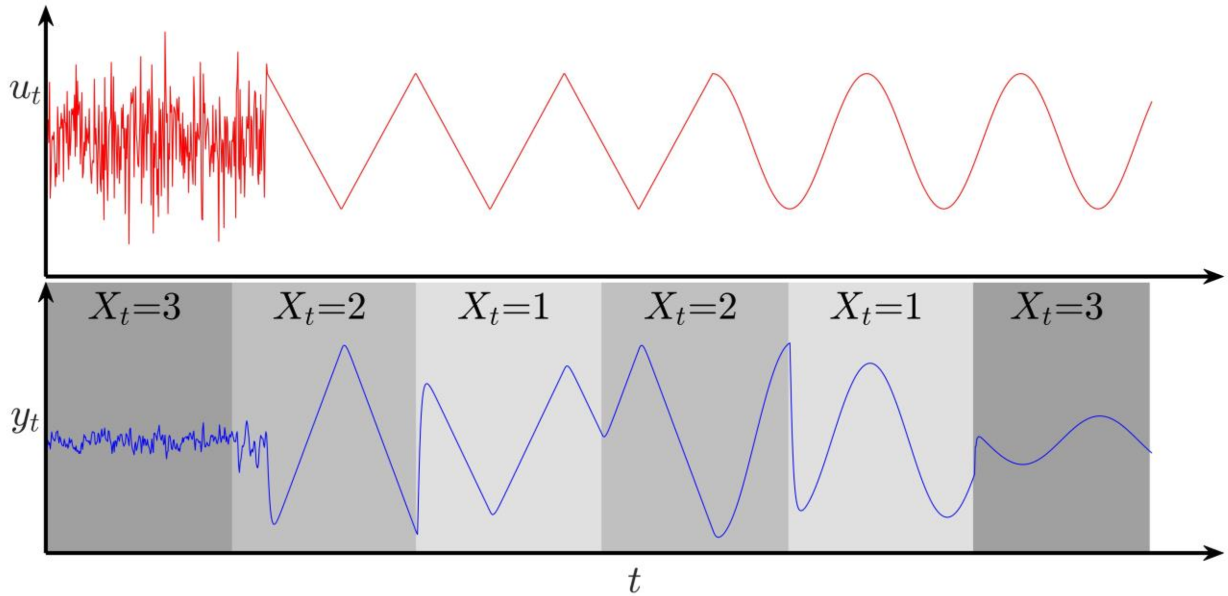


Markov Jump System

$$y_t = \mathbf{w}_{X_t}^\intercal \boldsymbol{\phi}_t + n_t$$

$$X_t \in \{1, 2, \ldots, n\}$$
$$\sim \text{MarkovChain}(\mathbf{P})$$

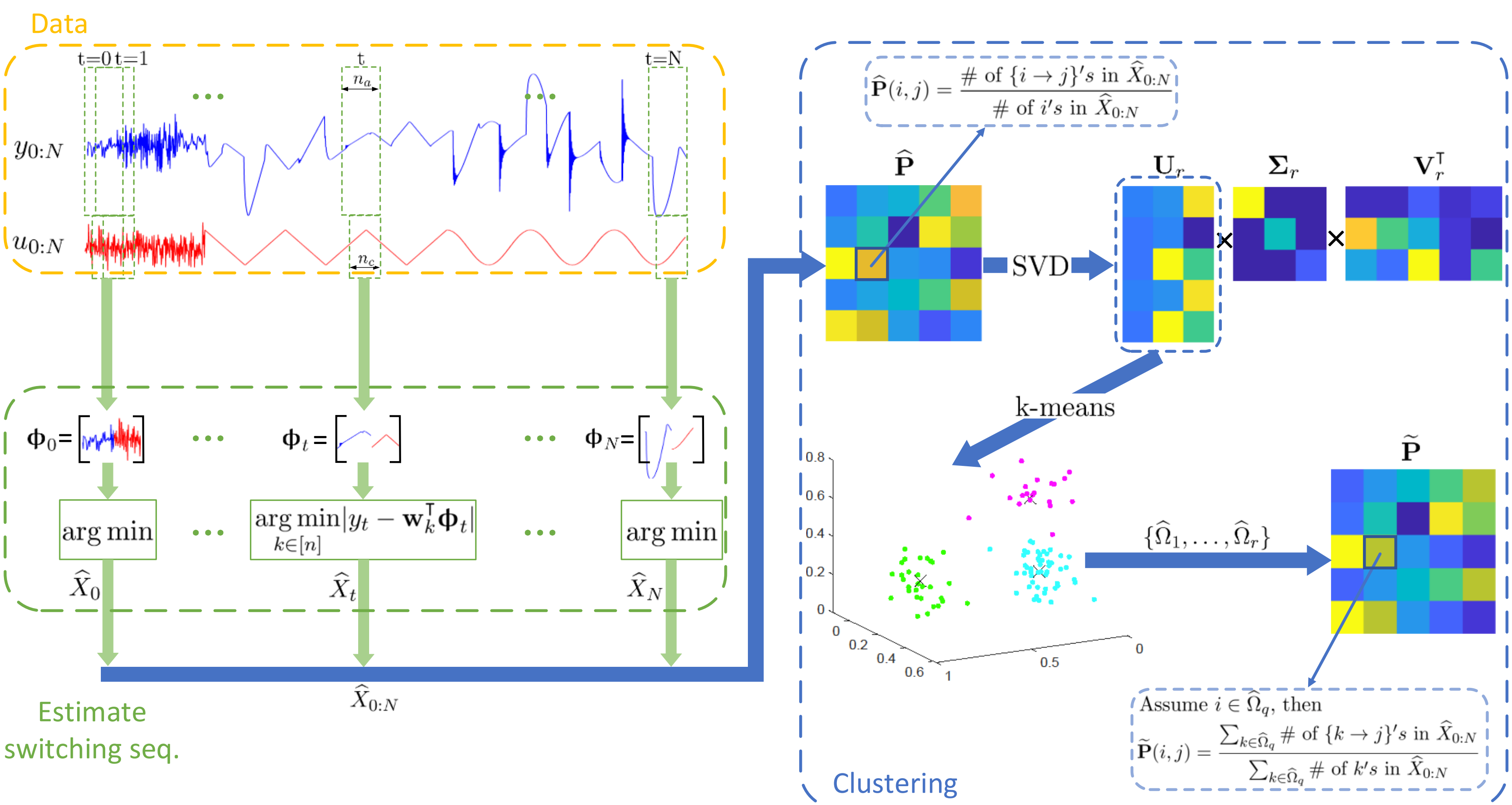
$$\boldsymbol{\phi}_t = \begin{bmatrix} y_{t-1} \\ \vdots \\ y_{t-n_a} \\ u_{t-1} \\ \vdots \\ u_{t-n_c} \end{bmatrix} \qquad \mathbf{w}_{X_t} = \begin{bmatrix} a_1(X_t) \\ \vdots \\ a_{n_a}(X_t) \\ c_1(X_t) \\ \vdots \\ c_{n_c}(X_t) \end{bmatrix}$$



Problem Formulation

- **Assume**
  - $$\mathbf{P} = \bar{\mathbf{P}} + \mathbf{\Delta}$$
    * $\bar{\mathbf{P}}$ — an underlying Markov matrix that is *r-aggregatable*, i.e. state space $\{1, 2, \ldots, n\}$ has partition $\{\Omega_1, \ldots, \Omega_r\}$ s.t. states within the same cluster share the same transition distributions. Mathematically, $\forall k, \forall i, j \in \Omega_k$
    $$\textbf{Prob}(X_{t+1}|X_t = i) = \textbf{Prob}(X_{t+1}|X_t = j)$$
    $$\text{or } \bar{\mathbf{P}}(i, :) = \bar{\mathbf{P}}(j, :).$$
    * $\mathbf{\Delta}$ — perturbation
  - Number of clusters r and dynamics $\{\mathbf{w}_k\}_{k=1}^{n}$ are known apriori
  - Disturbance $|n_t| < n_{\max}$
- **Goal**
  - Given trajectory $\{y_t, u_t\}_{t=0}^{N}$ only, estimate $\bar{\mathbf{P}}$ and its partition $\{\Omega_1, \ldots, \Omega_r\}$.

Our Approach



Theory

Notations

n	number of modes	τ_*	mixing time	N	trajectory length
r	number of clusters	$\sigma_i(\bar{\mathbf{P}})$	ith singular value		
$\|\boldsymbol{\Delta}\|$	perturbation	$\pi_{\max}, \pi_{\min}$	$\max, \min$ in $\boldsymbol{\pi}_{\text{stationary}}$		
		$\|\Omega_{(i)}\|$	size of ith largest cluster		

Definition 1 (*Clustering Error*(CE)).

$$CE(\widehat{\Omega}_{1:r}) = \frac{1}{n}\sum_{j=1}^{r} |\{i : i \in \Omega_j; i \notin \widehat{\Omega}_j\}|$$

Theorem 1. *When $\mathbf{P}$ is ergodic, $\|\boldsymbol{\Delta}\| \leq C_0 \frac{\sigma_r(\bar{\mathbf{P}})}{\sqrt{r}}\sqrt{1+\frac{|\Omega_{(r)}|}{|\Omega_{(1)}|}}$, and under some other mild conditions, for $\epsilon > 0$, with probability no less than $1-\exp(-C_1 N\epsilon^3/(\tau_*\pi_{\max}))$*

$$CE(\widehat{\Omega}_{1:r}) \leq C_2 \frac{nr}{|\Omega_{(r)}|}\left(\frac{\|\boldsymbol{\Delta}\|}{\sigma_r(\bar{\mathbf{P}})} + \epsilon\frac{\sigma_1(\bar{\mathbf{P}})}{\pi_{\min}\sigma_r(\bar{\mathbf{P}})}\right)^2$$

Theory

Adverse Factors				Favorable Factor	
n	number of modes	τ_*	mixing time		
r	number of clusters	$\sigma_1(\bar{\mathbf{P}})/\sigma_r(\bar{\mathbf{P}})$	condition number	N	trajectory length
$\|\boldsymbol{\Delta}\|$	perturbation	$\pi_{\max}/\pi_{\min}$	disparity in $\boldsymbol{\pi}_{\text{stationary}}$		
		$\|\Omega_{(1)}\|/\|\Omega_{(r)}\|$	disparity in cluster sizes		

Definition 1 (*Clustering Error*(CE))**.**

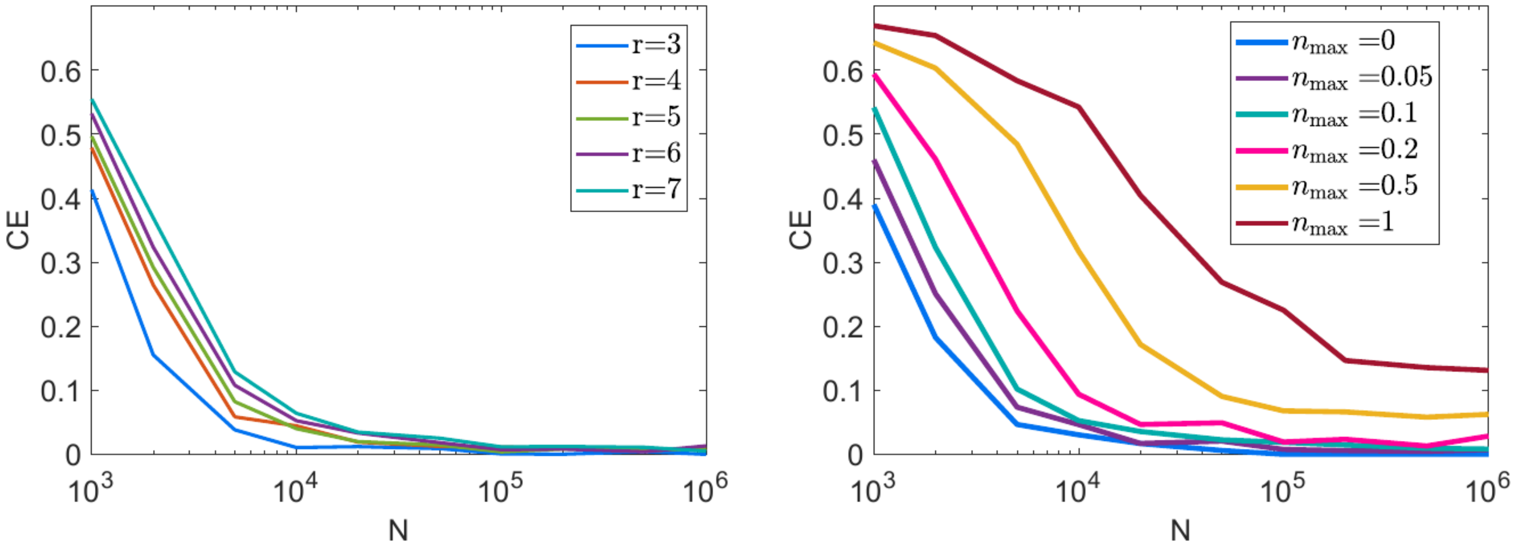
$$CE(\widehat{\Omega}_{1:r}) = \frac{1}{n}\sum_{j=1}^{r} |\{i : i \in \Omega_j; i \notin \widehat{\Omega}_j\}|$$

Theorem 1. *When* $\mathbf{P}$ *is ergodic,* $\|\boldsymbol{\Delta}\| \le C_0 \frac{\sigma_r(\bar{\mathbf{P}})}{\sqrt{r}} \sqrt{1 + \frac{|\Omega_{(r)}|}{|\Omega_{(1)}|}}$, *and under some other mild conditions, for* $\epsilon > 0$, *with probability no less than* $1 - \exp(-C_1 N \epsilon^3/(\tau_* \pi_{\max}))$

$$CE(\widehat{\Omega}_{1:r}) \le C_2 \frac{nr}{|\Omega_{(r)}|} \left(\frac{\|\boldsymbol{\Delta}\|}{\sigma_r(\bar{\mathbf{P}})} + \epsilon \frac{\sigma_1(\bar{\mathbf{P}})}{\pi_{\min}\sigma_r(\bar{\mathbf{P}})} \right)^2$$

Experiments

- $n_a{=}3, n_c{=}2, n{=}50$;
- Dynamics are generated by uniformly sampling poles on $(-1, 1)$;
- $u_t{\sim}\mathcal{N}(0, 1), n_t{\sim}Unif(-n_{\min}, n_{\max})$
- Partition $\Omega_{1:r}$ is sampled uniformly;
- $\bar{\mathbf{P}}(\Omega_k, :), \boldsymbol{\pi}_0 \sim DirichletDist$;
- $\boldsymbol{\Delta} = \mathbf{0}$ for now for simpler implementation.
- The errors are averaged over 100 experiments.



(a) CE vs N vs r, $n_{\max} = 0.1$ (b) CE vs N vs $n_{\max}$, $r = 6$

Experiments

- Mode clustering can help reduce the computation complexity

Compute Stationary Dist.			Linear Quadratic Regulator		
$r = 10$	$\mathbf{P}$	$\widetilde{\mathbf{P}}$	$r = 5$	Original MJS	Reduced MJS
$n = 1000$	0.029s	0.005s	$n = 20$	80.42s	3.04s
$n = 2000$	0.197s	0.025s	$n = 30$	168.34s	2.82s
$n = 5000$	1.318s	0.118s	$n = 40$	322.26s	2.85s
$n = 10000$	5.035s	0.426s	$n = 50$	483.47s	3.01s

Compute with the original one

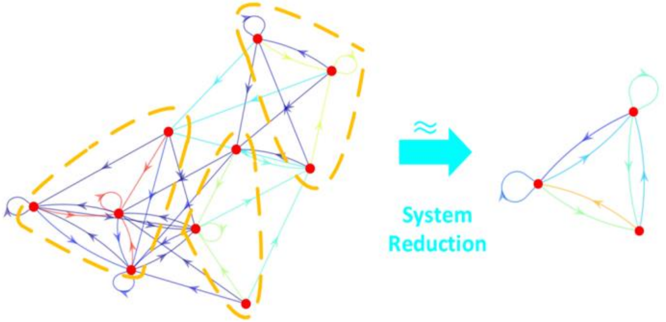
Compute with the surrogate one

Conclusion

- Summary
  - We propose a theoretically guaranteed approach to cluster modes of MJS based on their transition dynamics.
  - Simulation shows satisfactory performance in terms of clustering error.
  - This approach has prospect in system reduction.

Conclusion

- Summary
  - We propose a theoretically guaranteed approach to cluster modes of MJS based on their transition dynamics.
  - Simulation shows satisfactory performance in terms of clustering error.
  - This approach has prospect in system reduction.
- Future work



  - System reduction: combine the dynamics within the same cluster
  - Aggregatability $\rightarrow$ Lumpability