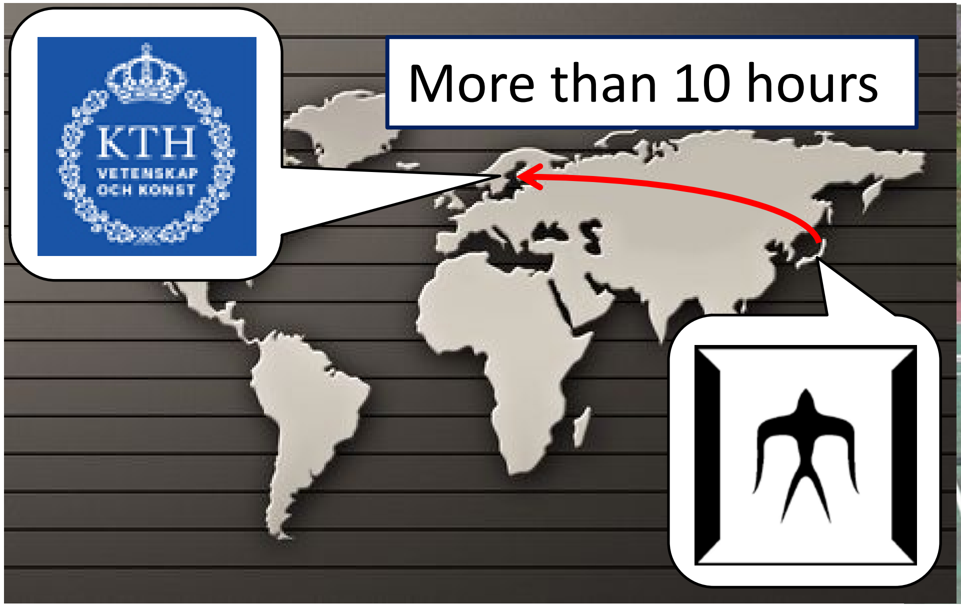
Clustering-based State Aggregation of Dynamical Networks

Takayuki Ishizaki

Ph.D. from Tokyo Institute of Technology (March 2012)

Research Fellow of the Japan Society for the Promotion of Science

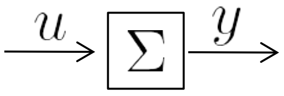


From Tokyo to Stockholm

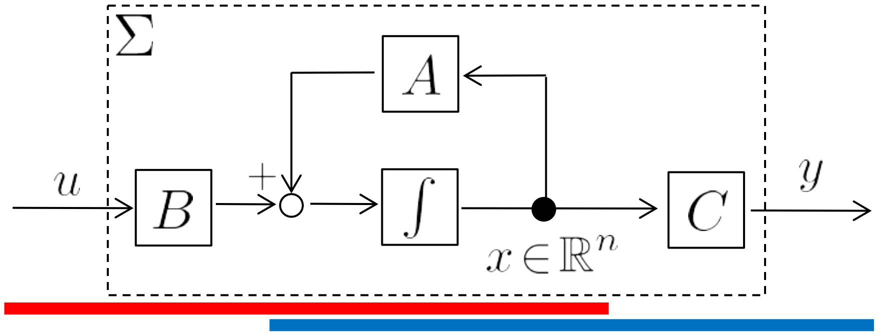
I am (was?) tennis man

Model Reduction via Projection

Given stable system



$$\Sigma : \begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$$



input-to-state map

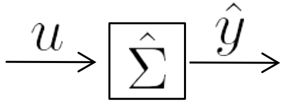
(A, B)

state-to-output map

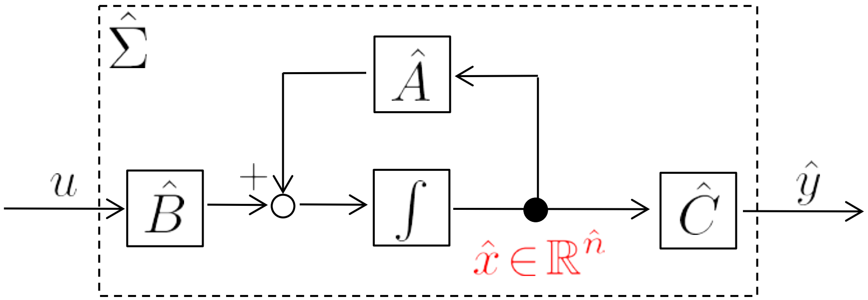
(A, C)

Find

Stable reduced model



$$\hat{\Sigma} : \begin{cases} \dot{\hat{x}} = \hat{A}\hat{x} + \hat{B}u \\ \hat{y} = \hat{C}\hat{x} \end{cases}$$



Dim. of state: $\hat{n} < n$

$Px = \hat{x},\ \ P \in \mathbb{R}^{\hat{n} \times n}$

✓ $PP^\dagger = I_{\hat{n}}$

$$(A, B, C) \longrightarrow (\hat{A}, \hat{B}, \hat{C}) = (PAP^\dagger, PB, CP^\dagger)$$

Find P such that $\|\Sigma - \hat{\Sigma}\|$ is small enough

Typical System Norms

✓ $\mathscr{L}[\cdot]$: Laplace transform
$\|\cdot\|_F$: Frobenius norm

Stable system $\Sigma : \begin{cases} \dot{x} = Ax + Bu \\ y = Cx \end{cases}$

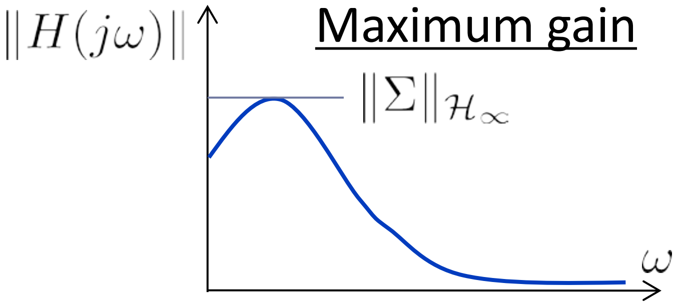
Solution $y(t) = \int_0^t h(t-\tau)u(\tau)d\tau$

Impulse response $h(t) := Ce^{At}B$

Transfer function $H(s) := \mathscr{L}[h] = C(sI_n - A)^{-1}B$

$\mathcal{H}_\infty$-norm

$$\|\Sigma\|_{\mathcal{H}_\infty} = \sup_{u \neq 0} \frac{\|y(t)\|_{\mathcal{L}_2}}{\|u(t)\|_{\mathcal{L}_2}} = \sup_{\omega \in \mathbb{R}} \|H(j\omega)\|$$



Maximum gain

$\mathcal{H}_2$ -norm

$$\|\Sigma\|_{\mathcal{H}_2} = \|h\|_{\mathcal{L}_2} = \left(\int_{-\infty}^{\infty} \|H(j\omega)\|_F^2 \frac{d\omega}{2\pi}\right)^{\frac{1}{2}}$$

Energy of $h(t) = Ce^{At}B$

✓ $\mathcal{L}_2$-norm of $f : \mathbb{R}_+ \to \mathbb{R}^{p \times m}$ $\quad \|f\|_{\mathcal{L}_2} := \left(\int_0^\infty \|f(t)\|_F^2 \, dt\right)^{\frac{1}{2}}$

Contents

- Clustering-based State Aggregation in terms of $\mathcal{H}_\infty$-norm
  - How to reduce systems while preserving network topology?
  - Use of positive tri-diagonalization
  - Application to diffusion process over complex network

- $\mathcal{H}_2$-aggregation of Positive Networks
  - Preservation of network topology as well as positivity
  - Use of controllability gramian
  - Application to Chemical Master Equation

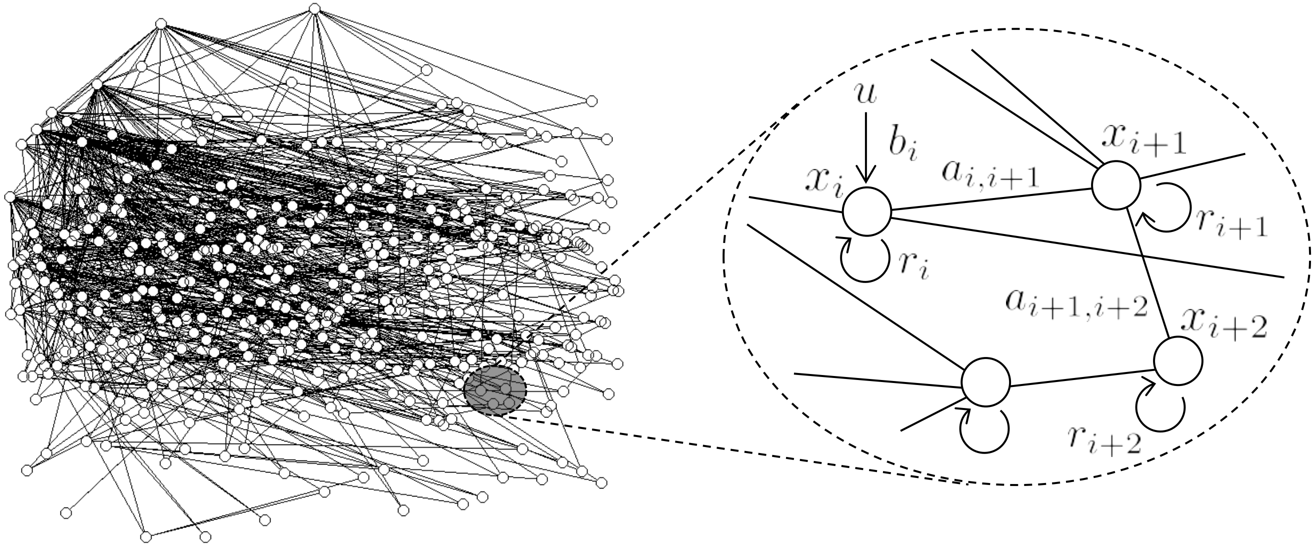
System Description

[Definition] **Bidirectional Network** (A, b)

$\dot{x} = Ax + bu$ with $A = \{a_{i,j}\} \in \mathbb{R}^{n \times n}$ and $b = \{b_i\} \in \mathbb{R}^n$

is said to be *bidirectional network* (A, b) if A is symmetric and stable.

Including reaction-diffusion systems: $\dot{x}_i = -r_i x_i + \sum_{j=1, j \neq i}^{n} a_{i,j} (x_j - x_i) + b_i u$

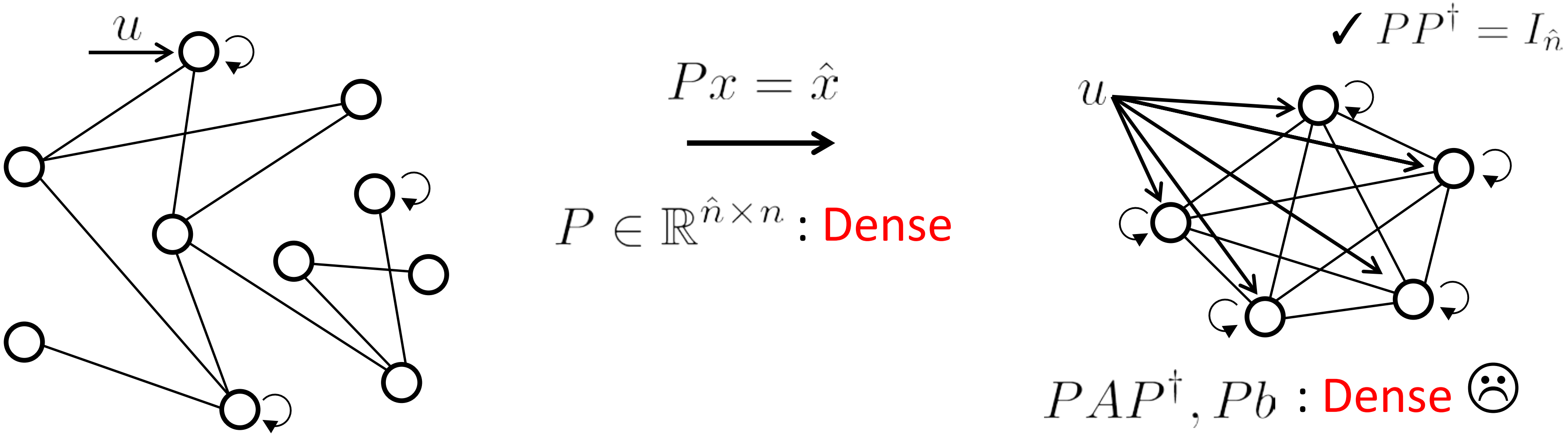


r_i : **reaction** of x_i $a_{i,j} = a_{j,i}$: **diffusion** between x_i and x_j

Traditional Model Reduction

- Traditional model reduction methods
  - Balanced truncation, Krylov projection, Hankel norm approximation
  - No specific structure in transformation matrix P

Drawback: Network structure (spatial information) is destroyed



Given (A, b) A, b : Sparse ☺

Reduced model $(PAP^\dagger, Pb)$

✓ $PP^\dagger = I_{\hat{n}}$

$$Px = \hat{x}$$

$$\longrightarrow$$

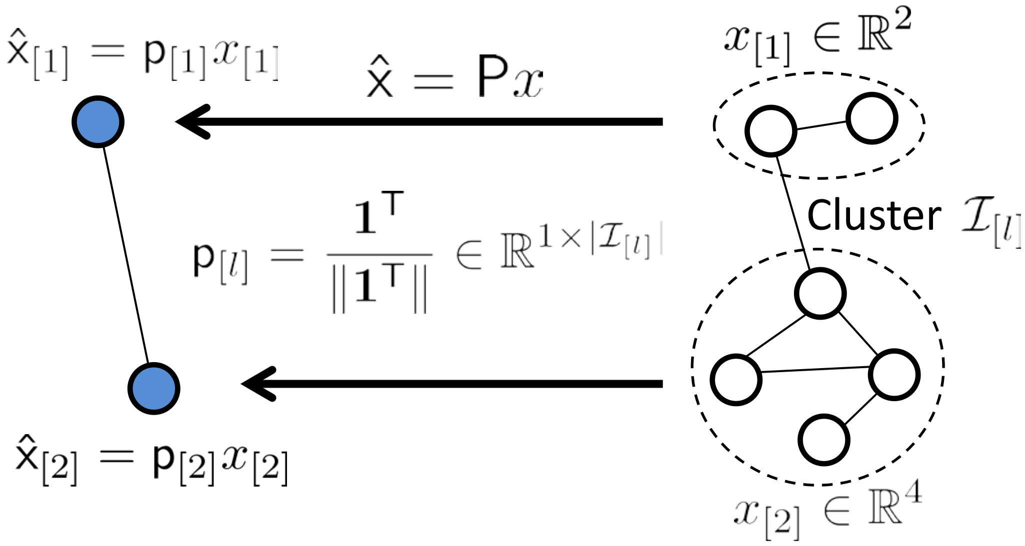
$P \in \mathbb{R}^{\hat{n} \times n}$: Dense

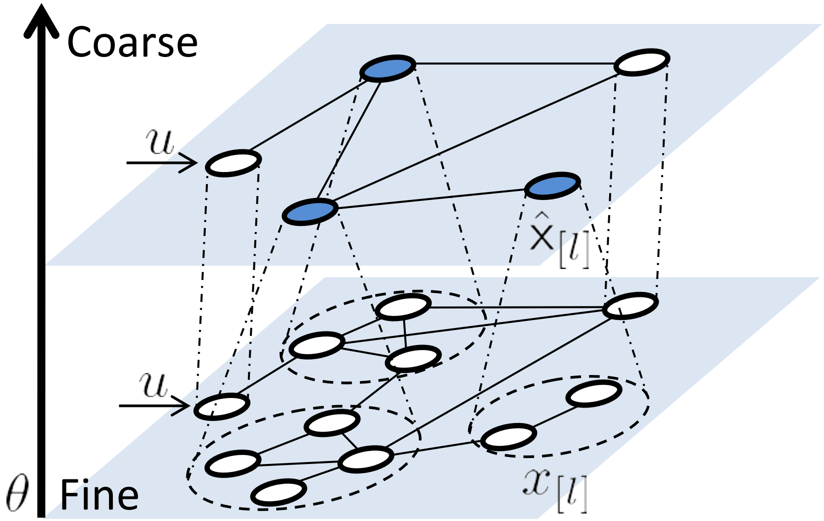
$PAP^\dagger, Pb$: Dense ☹

Need to impose suitable sparse structure on P

Clustering-based State Aggregation

- Aggregation of disjoint sets of states (clusters) $\{x_{[1]}, \ldots, x_{[L]}\}$
  - Block-diagonally structured aggregation matrix $\mathsf{P} = \mathrm{Diag}\left(\mathsf{p}_{[1]}, \ldots, \mathsf{p}_{[L]}\right)$
  - Interconnection topology among clusters is preserved ☺





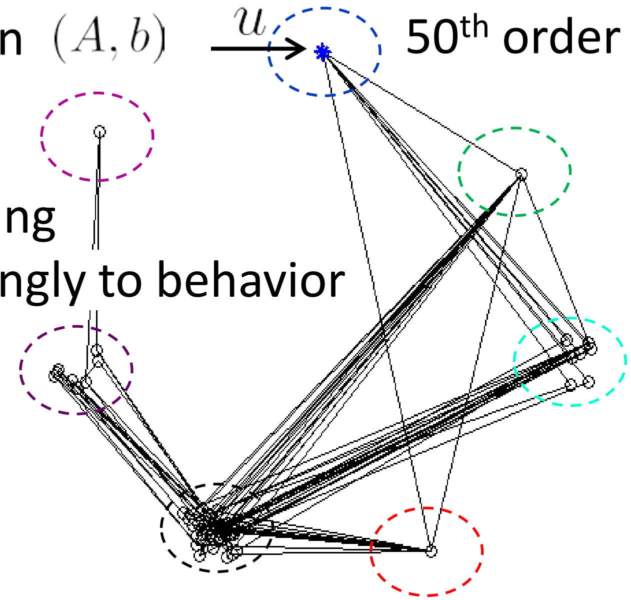
$$\hat{\mathsf{x}} = \mathsf{P} x \quad \Leftrightarrow \quad \begin{bmatrix} \hat{\mathsf{x}}_{[1]} \\ \hat{\mathsf{x}}_{[2]} \end{bmatrix} = \begin{bmatrix} \mathsf{p}_{[1]} & \\ & \mathsf{p}_{[2]} \end{bmatrix} \begin{bmatrix} x_{[1]} \\ x_{[2]} \end{bmatrix}$$

How to find reducible clusters?

✓ For simplicity, Aggregation = Averaging: $\mathbf{1}^\mathsf{T} = [1, \ldots, 1]$

Key Observation to Construct Reducible Clusters

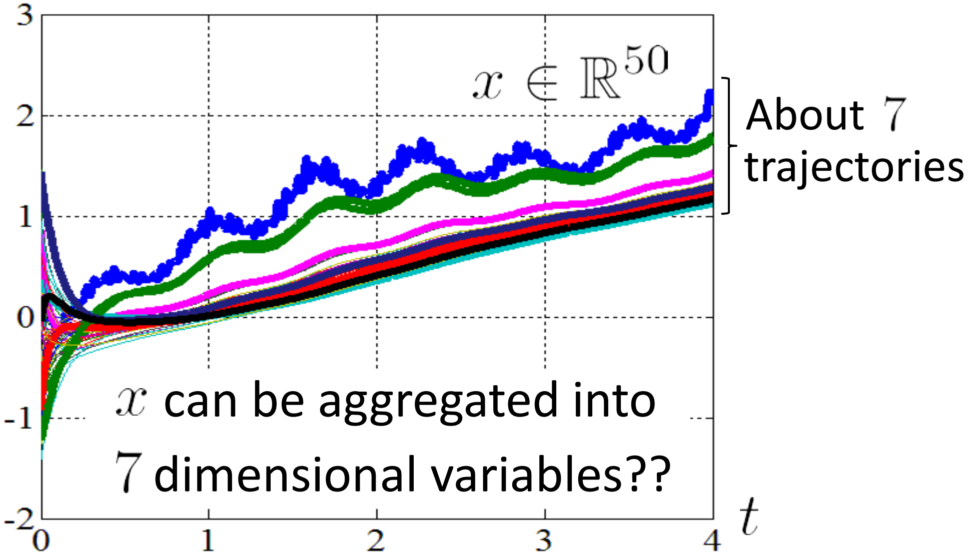
Given (A, b)



50th order

Clustering accordingly to behavior

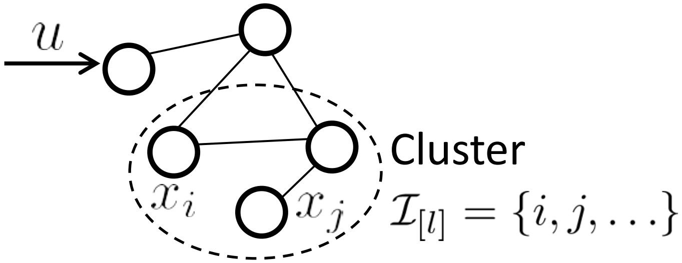
[State trajectory under random input]



About 7 trajectories

x can be aggregated into 7 dimensional variables??

[Definition] **Reducible Cluster**



A cluster $\mathcal{I}_{[l]}$ is said to be *reducible* if

$\forall i, j \in \mathcal{I}_{[l]}$ s.t. $x_i(t) \equiv x_j(t)$ under any input signal $u(t)$.

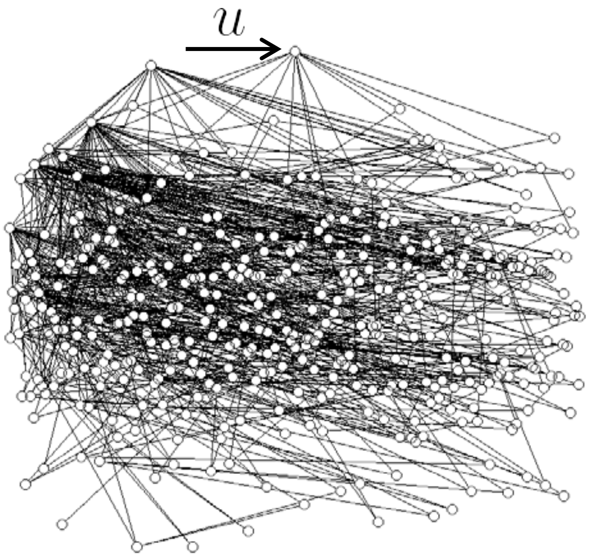
Positive Tri-diagonalization

[Lemma] For every bidirectional network (A, b), there exists a unitary $H \in \mathbb{R}^{n \times n}$ such that $(\tilde{A}, \tilde{b}) = (H^\mathsf{T} A H, H^\mathsf{T} b)$ has the structure below.

Bidirectional network (A, b)

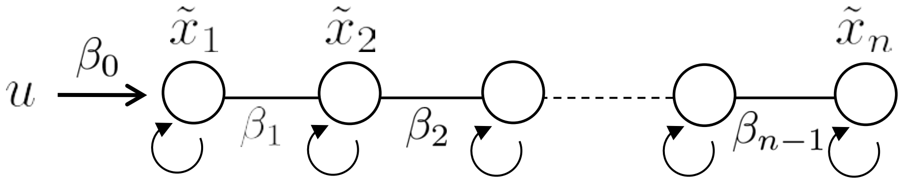
$$\dot{x} = Ax + bu, \quad A = A^\mathsf{T}$$

(not necessarily positive)



$$\overset{x = H\tilde{x}}{\longleftrightarrow}$$

Positive tri-diagonal realization $(\tilde{A}, \tilde{b})$



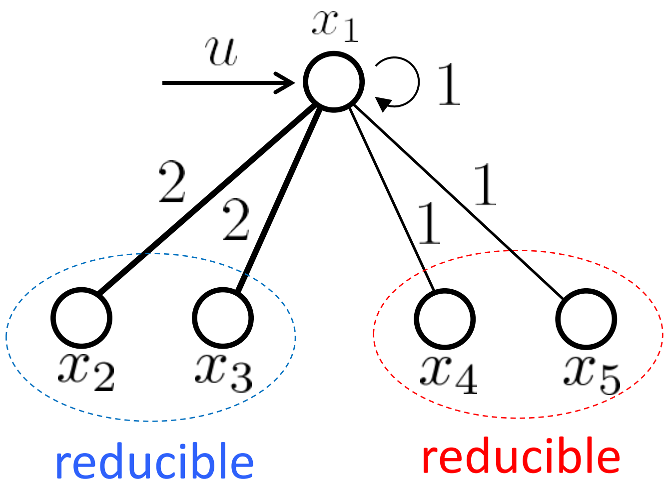
Metzler

$$\tilde{A} = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \beta_{n-1} \\ & & & \beta_{n-1} & \alpha_n \end{bmatrix} \quad \tilde{b} = \begin{bmatrix} \beta_0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\beta_i \geq 0$ for all $i \in \{0, \ldots, n-1\}$

Reducibility Characterization

Bidirectional network (A, b)



$(\tilde{A}, \tilde{b})$: positive tri-diagonal realization
H : transformation matrix

Index matrix

$$\Phi := H\mathrm{diag}(-\tilde{A}^{-1}\tilde{b})$$

$-\tilde{A}^{-1}\tilde{b}$: DC-gain ⇔ Maximal gain
due to positivity

$$\Phi = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 1.20 & -0.20 & 0 & 0 \\ \hline 0 & 1.20 & -0.20 & 0 & 0 \\ \hline 0 & 0.60 & 0.40 & 0 & 0 \\ \hline 0 & 0.60 & 0.40 & 0 & 0 \end{array}\right] \begin{array}{l} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{identical} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{identical} \end{array}$$

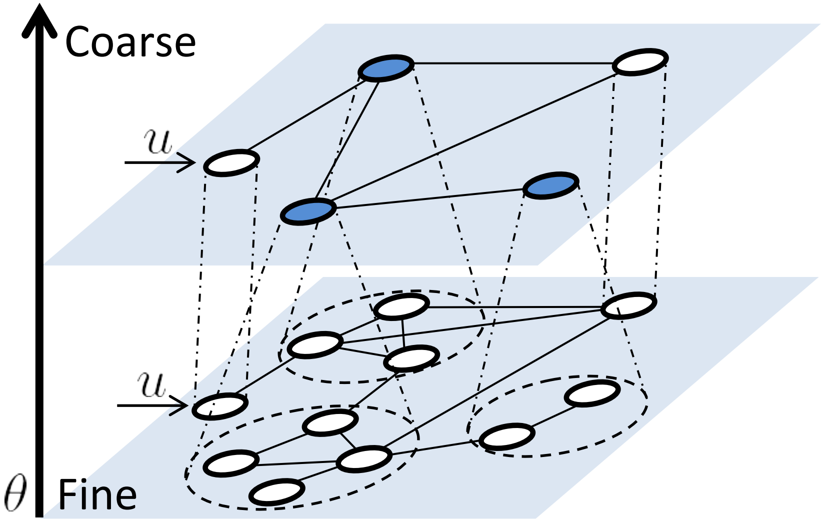
Cluster reducibility is characterized by rows of Φ

Reducible Cluster Aggregation

Reducibility: $\forall i,j \in \mathcal{I}_{[l]}$ s.t. $g_i(s) = g_j(s)$ $\quad\quad$ $\Phi = H\mathrm{diag}(-\tilde{A}^{-1}\tilde{b})$

[Theorem] A cluster $\mathcal{I}_{[l]}$ is reducible iff $\forall i,j \in \mathcal{I}_{[l]}$ s.t. $\mathrm{row}_i[\Phi] = \mathrm{row}_j[\Phi]$.

Furthermore, if all clusters are reducible, then $g(s) = \hat{\mathrm{g}}(s)$ holds.



Aggregated model $(\mathrm{P}A\mathrm{P}^{\mathsf{T}}, \mathrm{P}b)$

$$\hat{\mathrm{g}}(s) = \mathrm{P}^{\mathsf{T}}\left(sI_L - \mathrm{P}A\mathrm{P}^{\mathsf{T}}\right)^{-1}\mathrm{P}b$$

with $\mathrm{P} = \mathrm{Diag}(\mathrm{p}_{[1]}, \ldots, \mathrm{p}_{[L]}) \in \mathbb{R}^{L\times n}$

Dynamical network (A, b)

$$g\left(s\right) = \left(sI_n - A\right)^{-1} b$$

Relaxation to $\|\mathrm{row}_i[\Phi] - \mathrm{row}_j[\Phi]\| \leq \theta$??

Reducibility Relaxation

[Definition] θ**-reducible Cluster** ✓ $\|v\|_{l_\infty} = \|v^\mathsf{T}\|_{l_1}$ for **row** vector v

A cluster $\mathcal{I}_{[l]}$ is said to be *θ-reducible* if

$$\forall j \in \mathcal{I}_{[l]},\ \exists i \in \mathcal{I}_{[l]} \ \text{ s.t. } \ \|\mathrm{row}_i[\Phi] - \mathrm{row}_j[\Phi]\|_{l_\infty} \leq \theta.$$

θ : coarseness parameter

[Theorem] If all clusters are θ-reducible, then

$$\|g(s) - \hat{\mathrm{g}}(s)\|_{\mathcal{H}_\infty} \leq \sqrt{\alpha}\,\|(\mathsf{P}A\mathsf{P}^\mathsf{T})^{-1}\mathsf{P}A\|\theta$$

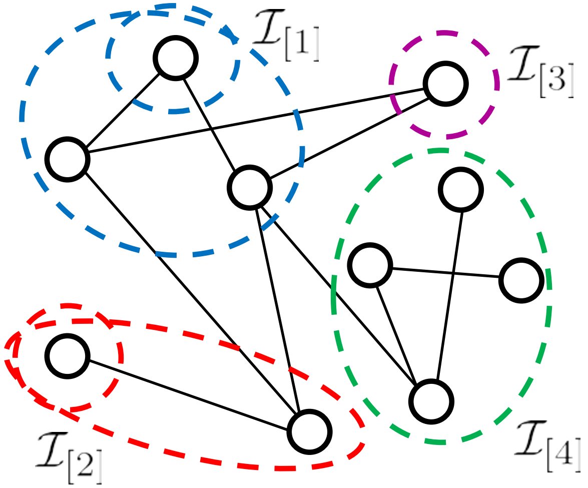
holds where $\alpha := \sum_{l=1}^{L} |\mathcal{I}_{[l]}|(|\mathcal{I}_{[l]}| - 1)$.

linear dependence on θ

Preservation: Stability and Interconnection topology among clusters ☺

In addition, $\hat{\mathrm{x}}_{[l]}$ represents average of original state $x_{[l]}$

Cluster Set Construction



- Give $\theta \in \mathbb{R}_+$, Initialize $\{\mathcal{I}_{[l]}\}_{l\in\mathbb{L}} = \emptyset,\ \mathbb{L} = \emptyset,\ l = 0$
  - While $\{\mathcal{I}_{[l]}\}_{l\in\mathbb{L}} \neq \{1,\ldots,n\}$
    - $l{+}{+},\ \mathbb{L} \leftarrow \{\mathbb{L}, l\}$
    - Choose $i \in \{1,\ldots,n\}\backslash\{\mathcal{I}_{[l]}\}_{l\in\mathbb{L}}$, Set $\mathcal{I}_{[l]} = \{i\}$
    - For all $j \in \{1,\ldots,n\}\backslash\{\mathcal{I}_{[l]}\}_{l\in\mathbb{L}}$,

      if (i,j) satisfies $(*)$, then $\mathcal{I}_{[l]} \leftarrow \{\mathcal{I}_{[l]}, j\}$

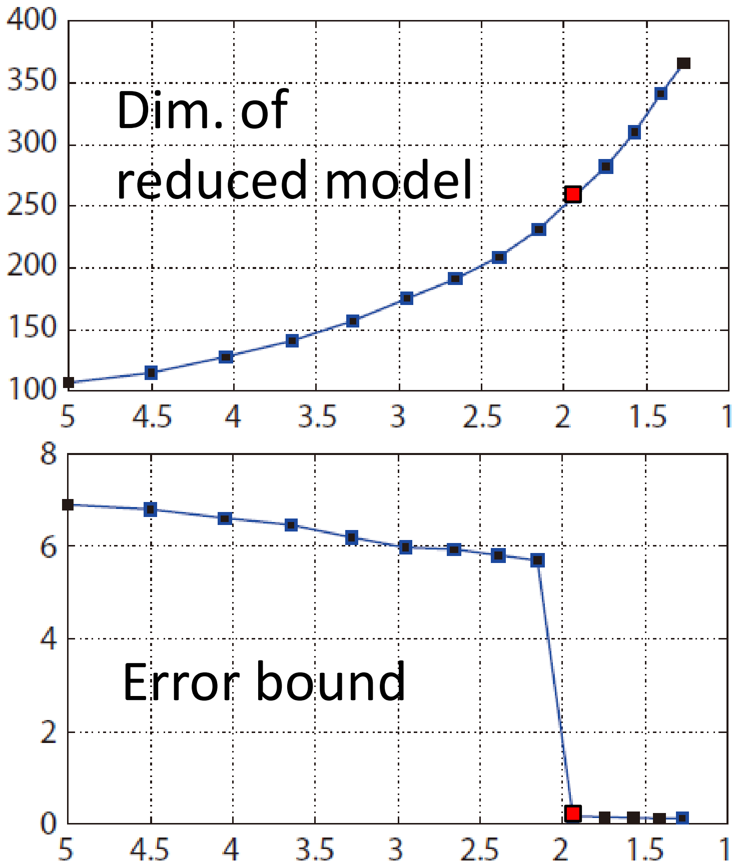
θ- reducibility : $(*)$

$$\|\mathrm{row}_i[\Phi] - \mathrm{row}_j[\Phi]\|_{l_\infty} \leq \theta \quad \text{where} \quad \Phi = H\mathrm{diag}(-\tilde{A}^{-1}\tilde{b})$$

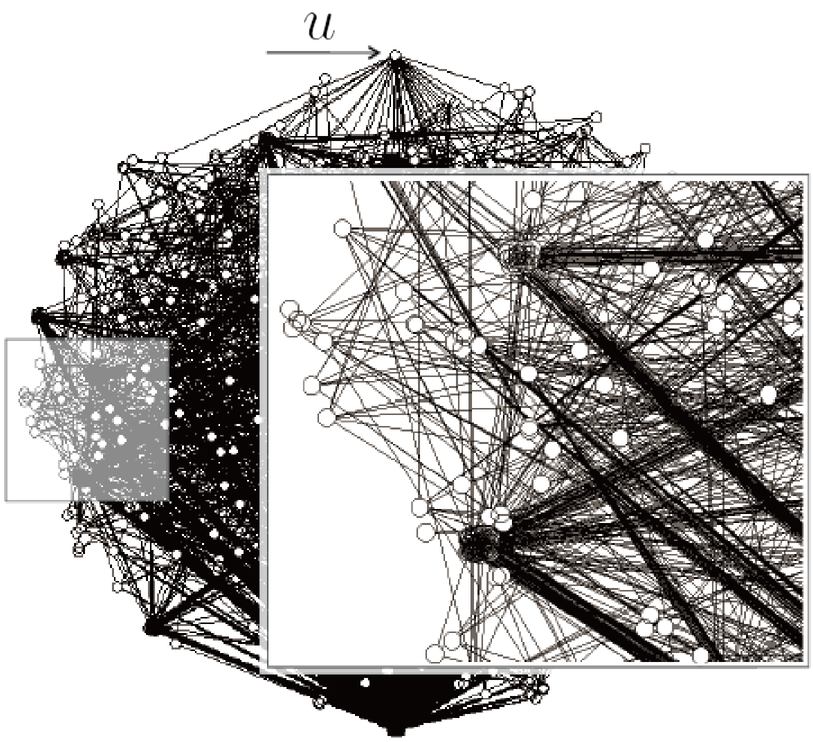
✓ Cluster set to be obtained is not necessarily unique

Numerical Example

- Diffusion process over the Holme-Kim model (3000 th dim.)
  - $\|g(s) - \hat{\mathrm{g}}(s)\|_{\mathcal{H}_\infty} \leq 0.16$ (less than 0.5% error) if $\theta = 1.82$



The value of θ



276 clusters

(276th dimensional model)

Contents

- Clustering-based State Aggregation in terms of $\mathcal{H}_\infty$-norm
  - How to reduce systems while preserving network topology?
  - Use of positive tri-diagonalization
  - Application to diffusion process over complex network

- $\mathcal{H}_2$-aggregation of Positive Networks
  - Preservation of network topology as well as positivity
  - Use of controllability gramian
  - Application to Chemical Master Equation

System Description

[Definition] **Positive Network** (A, b)

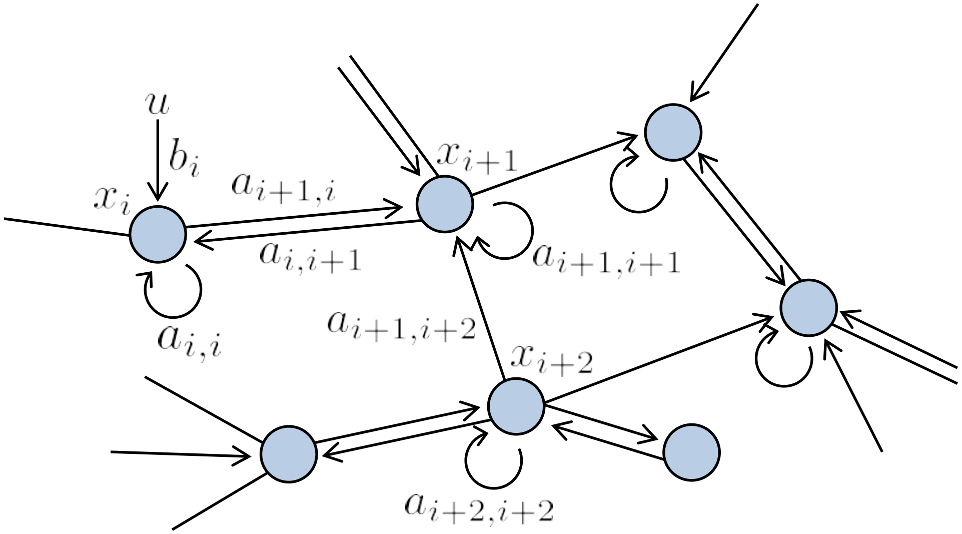
$\dot{x} = Ax + bu$ with $A = \{a_{i,j}\} \in \mathbb{R}^{n \times n}$ and $b = \{b_i\} \in \mathbb{R}^n$ is said to be *positive network* (A, b) if A is Metzler and (marginally) stable, and $b \in \mathbb{R}^n_+$.

Metzler matrix : having non-negative off-diagonal entries

$\mathbb{R}_+$: non-negative $\mathbb{R}$

non-negative property

$x(t) \in \mathbb{R}^n_+, \quad \forall u \in \mathbb{R}_+, \ \forall t \in [0, \infty)$

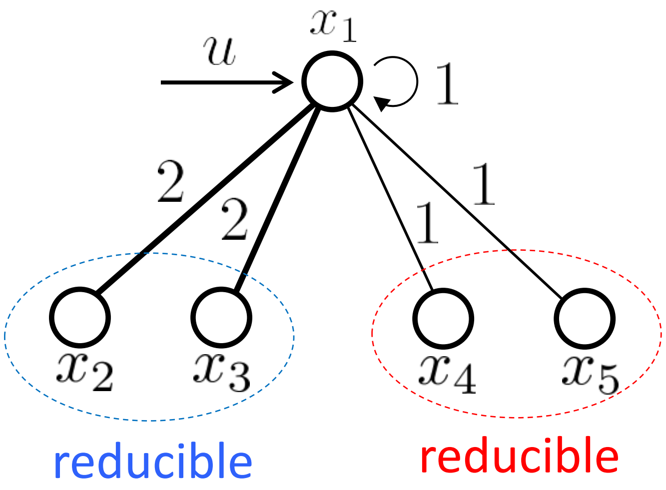


e.g., Heat diffusion systems, Electric circuit systems, Markovian processes

Model reduction while preserving positivity, stability and network

Reducibility Characterization ($\mathcal{H}_2$-case)

Given (A, b) with stable A



Controllability gramian

$$\Phi := \int_0^\infty e^{At} b \, (e^{At} b)^{\mathsf{T}} dt$$

✓ Lyapunov equation $A\Phi + \Phi A^{\mathsf{T}} + bb^{\mathsf{T}} = 0$

Cholesky factorization $\Phi_c \Phi_c^{\mathsf{T}} = \Phi$

$$\Phi_c = \left[\begin{array}{ccccc} 0.39 & 0 & 0 & 0 & 0 \\ \hline 0.21 & 0.19 & 0 & 0 & 0 \\ \hline 0.21 & 0.19 & 0 & 0 & 0 \\ \hline 0.24 & 0.18 & 0.05 & 0.00 & 0 \\ \hline 0.24 & 0.18 & 0.05 & 0.00 & 0 \end{array}\right] \begin{array}{l} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{identical} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \text{identical} \end{array}$$

Cluster reducibility is characterized by rows of Φ_c

$\mathcal{H}_2$-State Aggregation

[Definition] **θ-reducible Cluster** ✓ $\Phi_c\Phi_c^{\mathsf{T}} = \int_0^\infty e^{At}b\,(e^{At}b)^{\mathsf{T}}dt$

The cluster $\mathcal{I}_{[l]}$ is said to be *θ-weakly reducible* if

$$\forall j \in \mathcal{I}_{[l]},\ \ \exists i \in \mathcal{I}_{[l]}\ \ \text{s.t.}\ \ \|\mathrm{row}_i[\Phi_c] - \mathrm{row}_j[\Phi_c]\| \leq \theta.$$

θ : coarseness parameter

[Theorem] If all clusters are θ-weakly reducible, then

$$\|g(s) - \hat{\mathsf{g}}(s)\|_{\mathcal{H}_2} \leq \sqrt{\alpha}\,\|(sI_L - \mathsf{P}A\mathsf{P}^{\mathsf{T}})^{-1}\mathsf{P}A\|_{\mathcal{H}_\infty}\theta$$

holds where $\alpha := \sum_{l=1}^{L}|\mathcal{I}_{[l]}|(|\mathcal{I}_{[l]}| - 1)$.

linearly bounded by θ

Preservation: Stability, Positivity, Interconnection topology among clusters

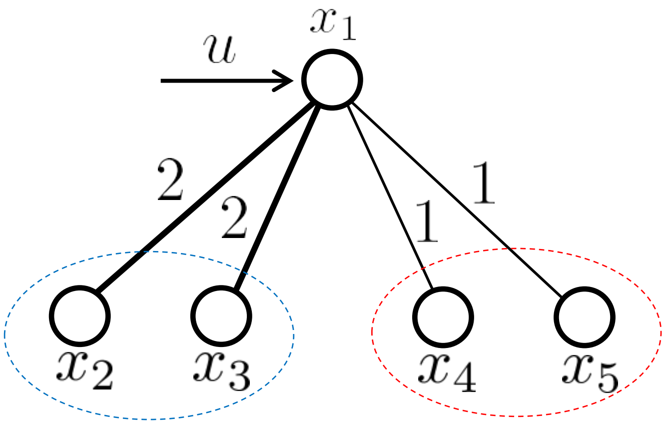
Generalization to Marginally Stable Positive Networks

Gramian is **not defined** if A has zero-eigenvalue ☹

Projected gramian $\Phi = \int_0^\infty W^\top e^{WAW^\top t} Wb\,(W^\top e^{WAW^\top t} Wb)^\top dt$

where $W \in \mathbb{R}^{(n-1)\times n}$ is orthogonal complement of v_l such that $v_l^\top A = 0$.

✓
- Controllability gramian of stable projected system $(WAW^\top, Wb)$
- Unique positive semi-definite matrix for (A, b)



$-A$: Graph Laplacian

Cholesky factorization $\Phi_c \Phi_c^\top = \Phi$

$$\Phi_c = \left[\begin{array}{cccc} -0.18 & -0.05 & 0 & 0 \\ \hline \color{blue}{0.07} & \color{blue}{-0.01} & \color{blue}{0} & \color{blue}{0} \\ \hline \color{blue}{0.07} & \color{blue}{-0.01} & \color{blue}{0} & \color{blue}{0} \\ \hline \color{red}{0.02} & \color{red}{0.04} & \color{red}{0} & \color{red}{0} \\ \hline \color{red}{0.02} & \color{red}{0.04} & \color{red}{0} & \color{red}{0} \end{array}\right] \begin{array}{l} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \color{blue}{\text{identical}} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} \color{red}{\text{identical}} \end{array}$$

Application to Chemical Master Equation（CME）

Michaelis-Menten system

$$S_1 + S_2 \underset{c_2}{\overset{c_1}{\rightleftarrows}} S_3$$

$$S_3 \xrightarrow{c_3} S_4 + S_2$$

c_i : reaction rate constant

ex) Initial number of S_1, S_2 are both $K = 2$

[Realizable distributions]

Realization probability

$$\begin{tikzcd} x_1(t)\begin{bmatrix}2\\2\\0\\0\end{bmatrix} \arrow[r, shift left] & x_2(t)\begin{bmatrix}1\\1\\1\\0\end{bmatrix} \arrow[l, shift left] \arrow[r, shift left] \arrow[d] & x_3(t)\begin{bmatrix}0\\0\\2\\0\end{bmatrix} \arrow[l, shift left] \arrow[d] \\ & x_4(t)\begin{bmatrix}1\\2\\0\\1\end{bmatrix} \arrow[r, shift left] & x_5(t)\begin{bmatrix}0\\1\\1\\1\end{bmatrix} \arrow[l, shift left] \arrow[d] \\ & & x_6(t)\begin{bmatrix}0\\2\\0\\2\end{bmatrix} \end{tikzcd}$$

State $x := [x_1, \ldots, x_n]^\mathsf{T}$ with $x_1(0) = 1$

CME expression: $\dot{x} = Ax, \quad x(0) = [1, 0, \ldots, 0]^\mathsf{T}$

- Continuous-time Markovian process
  - off-diagonal entries of A are non-negative
  - column sums of A are zero $\Leftrightarrow$ $\sum_{i=1}^{n} x_i(t) \equiv 1$ (zero-eigenvalue)
- $n = (K+1)(K+2)/2$ th dimensional

Numerical Example

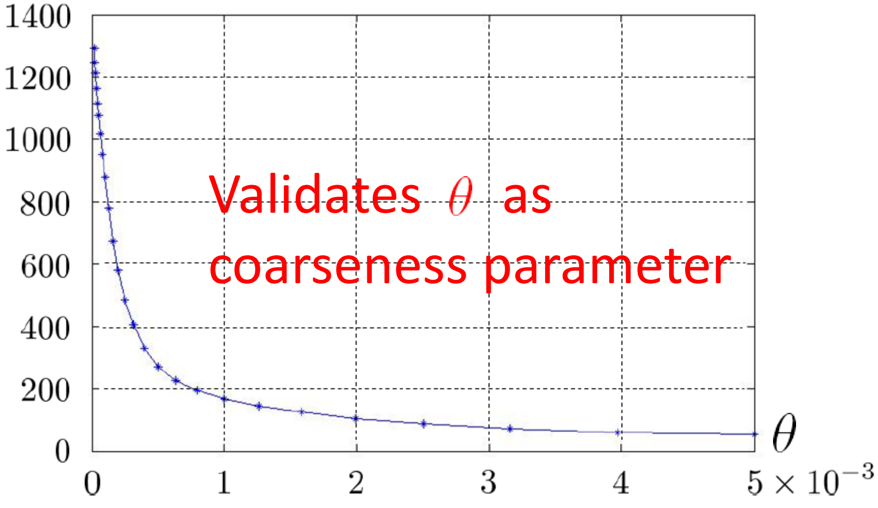
$$\dot{x} = Ax, \quad x(0) = [1, 0, \ldots, 0]^{\mathsf{T}} \xrightarrow{\mathsf{P}x = \hat{\mathsf{x}}} \dot{\hat{\mathsf{x}}} = \mathsf{P}A\mathsf{P}^{\mathsf{T}}\hat{\mathsf{x}}, \quad \hat{\mathsf{x}}(0) = \mathsf{P}x(0)$$

$n = 10011$ th order

if $\theta = 5 \times 10^{-5}$

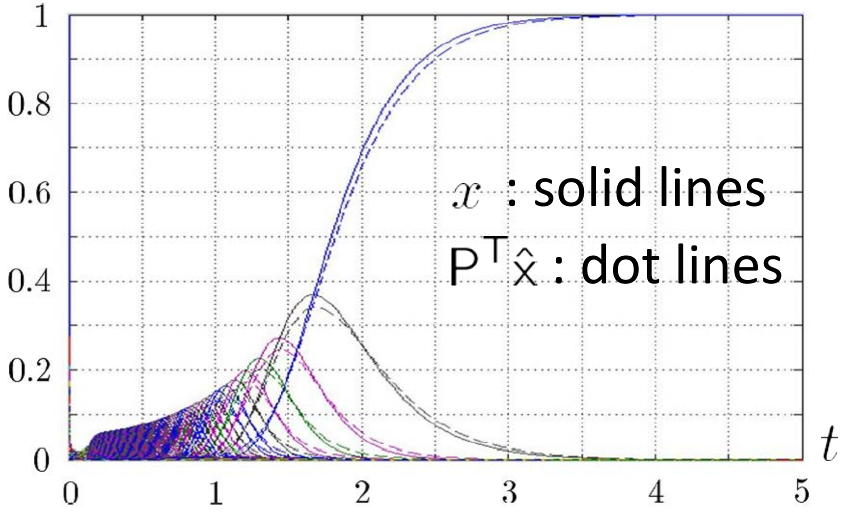
$L = 1077$ th order

[Reduced order versus θ]



Validates θ as coarseness parameter

[Trajectories of x and $\mathsf{P}^{\mathsf{T}}\hat{\mathsf{x}}$]



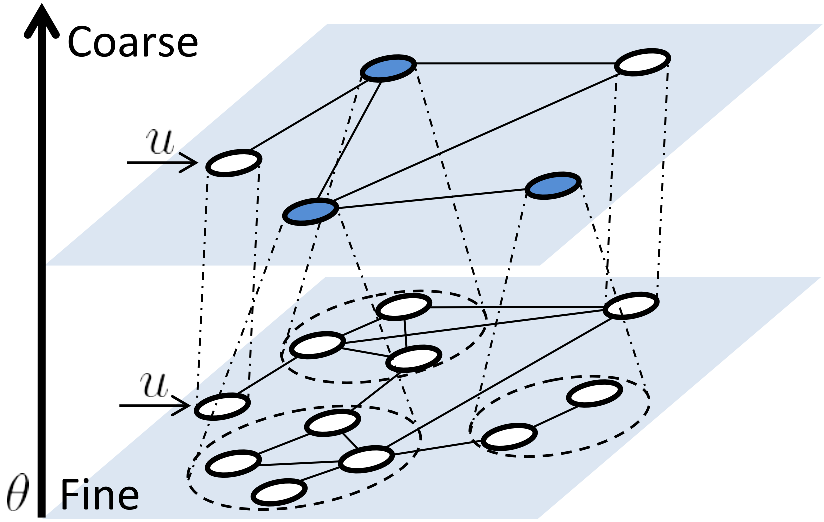
x : solid lines

$\mathsf{P}^{\mathsf{T}}\hat{\mathsf{x}}$: dot lines

Relative error of $x - \mathsf{P}^{\mathsf{T}}\hat{\mathsf{x}}$ in $\mathcal{H}_2$-norm: 2.4%

Summary

- Clustering-based State Aggregation
  - positive tri-diagonalization leads to $\mathcal{H}_\infty$-aggregation
  - controllability graman leads to $\mathcal{H}_2$-aggregation
- Preserving interconnection topology as well as stability, positivity
- Application to diffusion process over complex networks and CMEs



Aggregated model $(\mathsf{P}A\mathsf{P}^\mathsf{T}, \mathsf{P}b)$

$$\hat{\mathsf{g}}(s) = \mathsf{P}^\mathsf{T}\left(sI_L - \mathsf{P}A\mathsf{P}^\mathsf{T}\right)^{-1}\mathsf{P}b$$

with $\mathsf{P} = \mathrm{Diag}(\mathsf{p}_{[1]}, \ldots, \mathsf{p}_{[L]}) \in \mathbb{R}^{L \times n}$

Dynamical network (A, b)

$$g\left(s\right) = \left(sI_n - A\right)^{-1} b$$