Thrombin system reduction

Hoang M. Tran

November 14, 2014

Outline

Outline

Problem formulation

$$\dot{\boldsymbol{y}}(t) = \boldsymbol{f}(\boldsymbol{y}(t), \boldsymbol{k}) \qquad (1)$$

where

- $\boldsymbol{y} = (y_1, \ldots, y_n)$, $n = 34$, y_i: concentration of chemical species i
- $\boldsymbol{k}$: reaction rates, given constants
- $\boldsymbol{f}(\boldsymbol{y}(t), \boldsymbol{k}) = (f_1(\boldsymbol{y}(t), \boldsymbol{k}), \ldots, f_m(\boldsymbol{y}(t), \boldsymbol{k}))$, $m = 42$, there are 42 rate equation
- y_{14}: concentration of Thrombin

Problem formulation

Find another system of equations in terms of new variable $\boldsymbol{z}$ such that

- $\boldsymbol{z}(t)$ is a transformation of $\boldsymbol{y}(t)$
- $\dim(\boldsymbol{z}) \leq 4$

Outline

1. Problem formulation
2. $l_1 - \min$ formulation
3. Data driven system reduction method
   - Correlation analysis of time series from simulated human coagulation cascade dynamics
   - System reduction using linear transformation
4. Update week of November 14th 2014
   - Kuramoto model reconstruction: bound on total perturbation
   - Dimension reduction

$l_1 - \min$ formulation

$l_1 - \min$ formulation

$$\min |C|_1 \text{ s. t. } \begin{cases} ||\dot{\boldsymbol{z}}(t) - \boldsymbol{f}(\boldsymbol{z}(t), C)||_k & \leq \varepsilon_1 \forall t \in \{t_1, ..., t_N\} \\ ||z_{14} - f_{14}(\boldsymbol{z}(t), C)||_k & \leq \varepsilon_2 \end{cases}$$

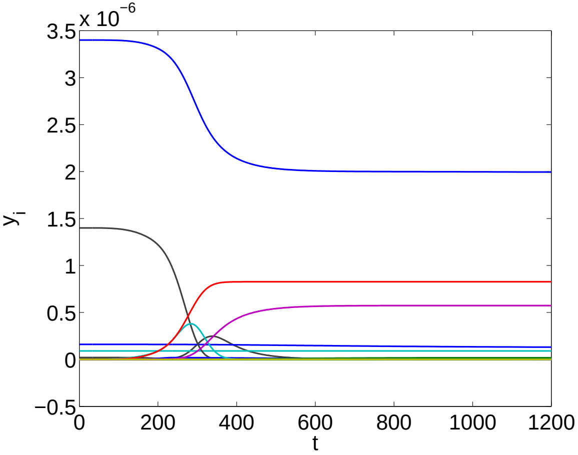
where$\varepsilon_1 > \varepsilon_2$

Note that the transformation: $\boldsymbol{y} \rightarrow \boldsymbol{z} = \boldsymbol{y}$.

Outline

1. Problem formulation
2. $l_1 -$ min formulation
3. Data driven system reduction method
   - Correlation analysis of time series from simulated human coagulation cascade dynamics
   - System reduction using linear transformation
4. Update week of November 14th 2014
   - Kuramoto model reconstruction: bound on total perturbation
   - Dimension reduction

Time series



Time series: observations

- Time series are higly correlated. $14 \to 25$: correlation about 1.
- Simulated data Y when solving(1) can be approximately represented in lower dimension space

SVD the simulated data $Y_{n\times N} \xrightarrow{A} Y_{n\times N} \approx A_{n\times p} Z_{p\times N}, p \ll n$, use A as transformation to reduce the dynamical system (1)

Dimension reduction using singular value decomposition

$$\begin{aligned} C_{N\times n} &= S_{N\times N}\Sigma_{N\times n}(V_{n\times n})^T \\ \tilde{C} &= Z_{N\times p}\Sigma_{p\times p}(V_{p\times n})^T \\ &\approx C \\ \tilde{C}^T &= (\Sigma_{p\times p}(V_{p\times n})^T)^T S_{N\times p}^T \end{aligned} \tag{2}$$

- N: sample size, p: reduced dimension.

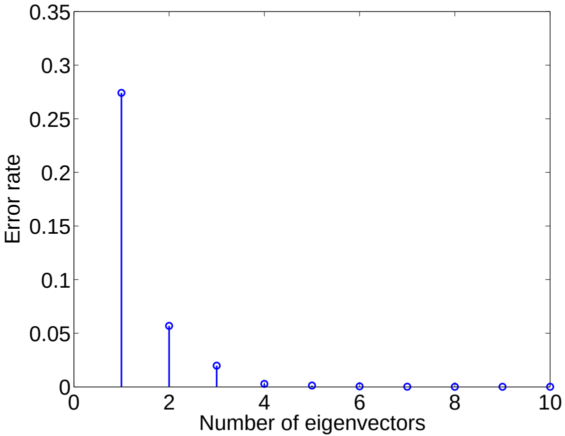
Apply (2) to $C = Y_{n\times N}^T$, denote $Z_{p\times N} = S_{N\times p}^T$,
$A = (\Sigma_{p\times p}(V_{p\times n})^T)^T = V_{p\times n}\Sigma_{p\times p}^T$

$$\tilde{Y}_{n\times N} = AZ_{p\times N} \tag{3}$$

$$\mathbf{y}(t) = A\mathbf{z}(t) \tag{4}$$

$$\mathbf{z}(t) = A^{-1}\mathbf{y}(t) \tag{5}$$

Dimension reduction using singular value decomposition



Error rate: $\frac{||\tilde{Y}^T - Y||_F}{||Y||_F}$

Dimension reduction using linear transformations: singular value decomposition

$$
\begin{aligned}
\boldsymbol{y}(t) &\approx A\boldsymbol{z}(t)\\
\boldsymbol{z}(t) &\approx A^{-1}\boldsymbol{y}(t)\\
\frac{d\boldsymbol{z}(t)}{dt} &= A^{-1}\frac{d\boldsymbol{y}(t)}{dt}\\
\frac{d\boldsymbol{z}(t)}{dt} &= A^{-1}\boldsymbol{f}(\boldsymbol{y}(t),\boldsymbol{k})\\
\frac{d\boldsymbol{z}(t)}{dt} &\approx A^{-1}\boldsymbol{f}(A\boldsymbol{z}(t),\boldsymbol{k}) \qquad (6)
\end{aligned}
$$

- A^{-1}: may need to use pseudo inverse.
- From $\boldsymbol{z}(t)$ we can get Thrombin and other components by $\boldsymbol{y}(t) = A\boldsymbol{z}(t)$.
- (6): equation for defining new variable z

Validation procedure

1. Solve the original system numerically to get $Y_{N\times n}$, $N = 1200, n = 34$, **not 29**.
2. Do SVD $Y_{N\times n} = Z_{N\times N}\Sigma_{N\times n}(V_{n\times n})^T$, $A = (\Sigma_{p\times p}(V_{p\times n})^T)^T$, $\tilde{Y}^{svd} = (\Sigma_{p\times p}(V_{p\times n})^T)^T Z^T_{N\times p}$
3. Solve $\dot{\boldsymbol{z}}(t) \approx A^{-1}\boldsymbol{f}(A\boldsymbol{z}(t), \boldsymbol{k})$ to get $(Z^{reduced}_{N\times p})^T$
4. Compute $(\tilde{Y}^{reconstruct})^T = A(Z^{reduced}_{N\times p})^T$
5. Compute error rate using $\tilde{Y}^{reconstruct}$, $\tilde{Y}^{svd}_{N\times N}$

Result - observations

(1)	1	2	3	4	5	6	7	8	9	10
(2)	-	stiff	-	stiff	stiff	-	stiff	stiff	stiff	stiff
(3)	0.86	5.78	0.17	18	108	**0.06**	**0.07**	6.68	5.57	1.8

Row (1) - number of eigenvectors, p or $\dim(\boldsymbol{z})$

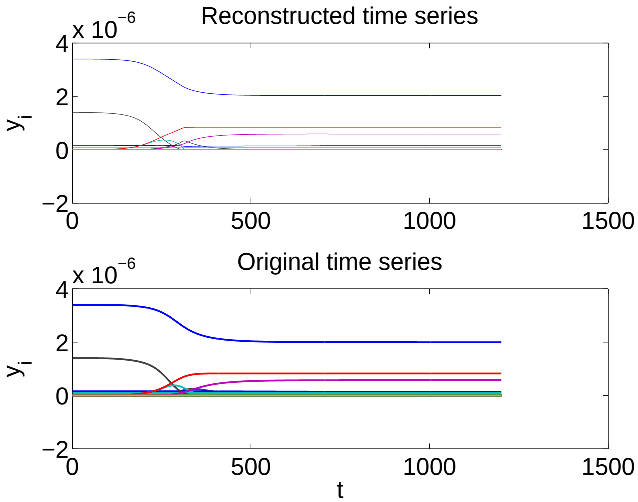
Row (2) - nonstiffness of the reconstructed system

$\dot{\boldsymbol{z}}(t) \approx A^{-1}\boldsymbol{f}(A\boldsymbol{z}(t), \boldsymbol{k})$

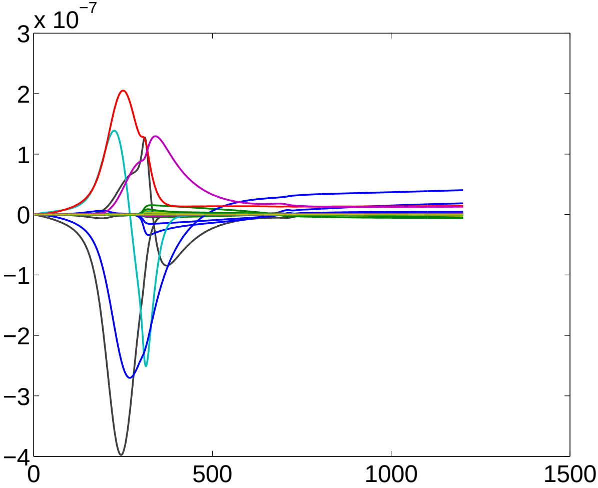
Rows (3): error rate computed using $\tilde{Y}^{reconstruc}$

- Training data length does matter
- The new system might become very stiff
- Accuracy does not linearly depends on training data length
- $p = 3$ or $p = 6$

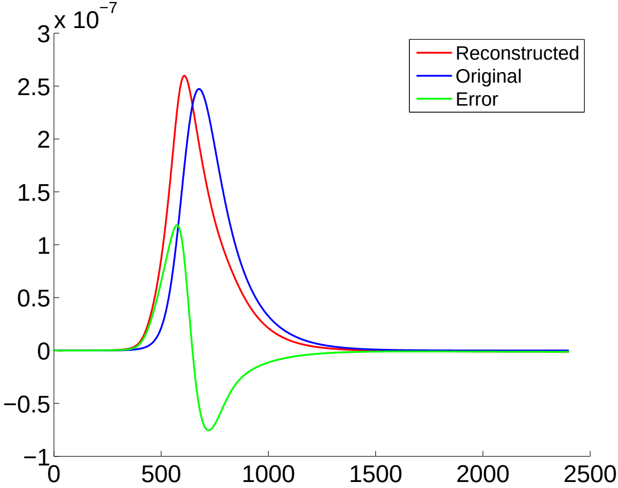
Time series comparision $p = 6$



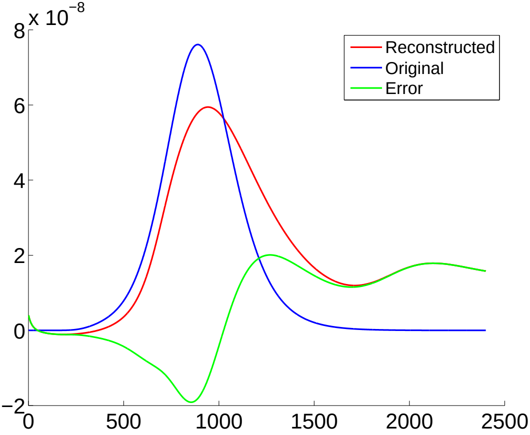
Time series comparision - error



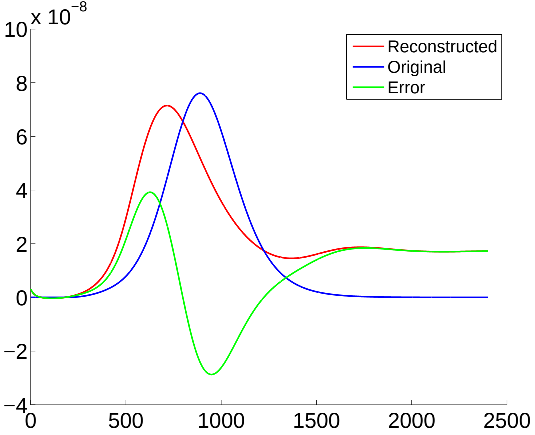
Thrombin time series



Thrombin time series - using the one initial condition, use another initial condition for validation



Thrombin time series - using same initial conditions



Result - some observations

- Using higher dimension space to represent new variable does not guarantee smaller error
- The approximated system might become very stiff
- A depends on intitial condition in the original system

$\Rightarrow$need a mathematical formulation and perturbation analysis

Outline

1. Problem formulation
2. $l_1 - \min$ formulation
3. Data driven system reduction method
   - Correlation analysis of time series from simulated human coagulation cascade dynamics
   - System reduction using linear transformation
4. Update week of November 14th 2014
   - Kuramoto model reconstruction: bound on total perturbation
   - Dimension reduction

Kuramoto model reconstruction: Bound on total perturbation

Theorem

For matrices R, R^0 if $||R-I||_F + ||\Delta R||_F < 1$ we have:

$$||R^{-1} - (R^0)^{-1}||_F \leq ||\Delta R||_F \frac{1}{(1 - ||R-I||_F - ||\Delta R||_F)^2}$$

where $\Delta R = R - R^0$.

Proof.

Apply the Theorem 6.2.30 from Johnson's book for:
$f(t) = \sum_{k=0}^{\infty}(-1)^k t^k = \frac{1}{1+t}$, $f_{abs}(t) = \sum_{k=0}^{\infty} t^k = \frac{1}{1-t}$, we have
$f_{abs}^{'}(t) = \frac{1}{(1-t)^2}$, $f(R-I) = (I+R-I)^{-1} = R^{-1}$,
$f(R^0 - I) = (R^0)^{-1}$

$$\begin{aligned} ||R^{-1} - (R^0)^{-1}||_F &= ||f(R-I) - f(R^0 - I)||_F \\ &\leq ||R - R^0||_F f_{abs}^{'}(||R-I||_F + ||(R-R^0)||_F) \end{aligned}$$

Kuramoto model reconstruction: Bound on total perturbation

Theorem

For the network untangling method, $P^0 = S^0 R^0$, the total perturbation $P - S^0 R = -(\Delta S)R$ is bounded by

$$(||P||_F + ||\Delta P||_F)||\Delta R||_F \frac{1}{(1 - ||R - I||_F - ||\Delta R||_F)^2} ||R||_F + ||\Delta P||_F$$

Proof.

$$\begin{aligned}
P^0 &= S^0 R^0 \Rightarrow S^0 = P^0 (R^0)^{-1} && (7)\\
S &= PR^{-1} \Rightarrow \Delta S = PR^{-1} - P^0 (R^0)^{-1} && (8)\\
\Delta S &= (P^0 + \Delta P)R^{-1} - P^0 (R^0)^{-1} && (9)\\
&= P^0 (R^{-1} - (R^0)^{-1}) + \Delta P R^{-1} && (10)\\
&= (P - \Delta P)(R^{-1} - (R^0)^{-1}) + \Delta P R^{-1} && (11)\\
(\Delta S)R &= (P - \Delta P)(R^{-1} - (R^0)^{-1})R + \Delta P && (12)
\end{aligned}$$

Problem with SVD based model reduction

SVD the simulated data $Y_{n\times N} \xrightarrow{A} Y_{n\times N} \approx A_{n\times p} Z_{p\times N}, p \ll n$, use A as transformation to reduce the dynamical system (1)

- $Y_{N\times n} = Z_{N\times N}\Sigma_{N\times n}(V_{n\times n})^T$, $A = (\Sigma_{p\times p}(V_{p\times n})^T)^T$, $\boldsymbol{y}(t) = A\boldsymbol{z}(t) + \varepsilon_1(t)$
- $\dot{\boldsymbol{z}}(t) \approx A^{-1}\boldsymbol{f}(A\boldsymbol{z}(t), \boldsymbol{k})$ to get $(Z^{reduced}_{N\times p})^T$
- Compute $(\tilde{Y}^{reconstruct})^T = A(Z^{reduced}_{N\times p})^T$

Problem: Not robust to the number of component used: stiffness issue, big error rate.

Dimension reduction using linear transformations

A, B are linear transformations that we need to define

$$
\begin{aligned}
\boldsymbol{y}(t) &= A\boldsymbol{z}(t) + \varepsilon_1(t) \\
\boldsymbol{z}(t) &= B\boldsymbol{y}(t) - \varepsilon_2(t)) \\
\dot{\boldsymbol{z}}(t) &= B\dot{\boldsymbol{y}}(t) - \dot{\varepsilon}_2(t)) \\
&= B\boldsymbol{f}(\boldsymbol{y}(t), \boldsymbol{k}) - \dot{\varepsilon}_2(t) \\
&= B\boldsymbol{f}(A\boldsymbol{z}(t) + \varepsilon_1(t), \boldsymbol{k}) - \dot{\varepsilon}_2(t) \\
\dot{\boldsymbol{z}}(t) - B\boldsymbol{f}(A\boldsymbol{z}(t), \boldsymbol{k}) &= B\boldsymbol{f}(A\boldsymbol{z}(t) + \varepsilon_1(t), \boldsymbol{k}) - \dot{\varepsilon}_2(t) - B\boldsymbol{f}(A\boldsymbol{z}(t), \boldsymbol{k}) \\
&= B\left[\boldsymbol{f}(A\boldsymbol{z}(t) + \varepsilon_1(t), \boldsymbol{k}) - \boldsymbol{f}(A\boldsymbol{z}(t), \boldsymbol{k})\right] - \dot{\varepsilon}_2(t) \\
&\approx B\frac{J(\boldsymbol{f})}{J(\boldsymbol{y})}(A\boldsymbol{z}(t) + \varepsilon_1(t), \boldsymbol{k})\varepsilon_1(t) - \dot{\varepsilon}_2(t) + o(||\varepsilon_1(t)|| \\
&\quad \text{(linear approximation at } A\boldsymbol{z}(t) + \varepsilon_1(t), \text{ not } A\boldsymbol{z}(t))
\end{aligned}
$$

Dimension reduction using linear transformations

$$\dot{\boldsymbol{z}}(t) - B\boldsymbol{f}(A\boldsymbol{z}(t), \boldsymbol{k}) \approx \boldsymbol{e}(t)$$
$$\boldsymbol{e}(t) = B\frac{J(\boldsymbol{f})}{J(\boldsymbol{y})}(\boldsymbol{y}(t), \boldsymbol{k})\varepsilon_1(t) - \dot{\varepsilon_2}(t) + o(||\varepsilon_1(t)||) \tag{13}$$

We need to find A and B such that

- $A\boldsymbol{z}(t) \approx \boldsymbol{y}(t)$. A is estimated from SVD of simulated data on $\boldsymbol{y}(t)$: $Y_{N\times n} = Z_{N\times N}\Sigma_{N\times n}(V_{n\times n})^T$, $A = (\Sigma_{p\times p}(V_{p\times n})^T)^T$,
- $||\boldsymbol{e}(t)||$ is as small as possible.
    - We can assume that $\varepsilon_2(t) = 0$
    - Current approach: $B = pinv(A)$ (Moore-Penrose pseudoinverse). Some row maybe very big.
    - Another option is choosing $B_{N\times p}$ such that $||B\frac{J(\boldsymbol{f})}{J(\boldsymbol{y})}(\boldsymbol{y}(t), \boldsymbol{k})||$ is minimized. Need to make sure that none of the rows of B is very big.