Subspace Identification and Spectral Learning of Dynamical Systems

Jun.21, 2018

Takehisa YAIRI（矢入健久）
E-mail: yairi@ailab.t.u-tokyo.ac.jp

Outline

- What is subspace identification ?
- Least square linear regression and singular value decomposition
- Deterministic and stochastic subspace identification
  - Extended state space representation
  - Canonical correlation analysis
- Spectral learning of dynamical systems

(Review) Learning LDS

- Model: $$\begin{cases} \boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{B}\boldsymbol{u}_t + \boldsymbol{v} & \boldsymbol{v} \sim N(\boldsymbol{0}, \boldsymbol{Q}) \\ \boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{w} & \boldsymbol{w} \sim N(\boldsymbol{0}, \boldsymbol{R}) \end{cases}$$
- Given :
  - Observation sequence : $\boldsymbol{y}_{1:T}$
  - Input sequence : $\boldsymbol{u}_{1:T}$
- Find :
  - System matrices : $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C},$
  - Noise covariance matrices: $\boldsymbol{Q}, \boldsymbol{R}$
  - Initial state distribution: $p(\boldsymbol{x}_1) \sim N(\boldsymbol{m}_0, \boldsymbol{V}_0)$
  - State sequence: $\boldsymbol{x}_{1:\mathrm{T}}$

(Review) EM Algorithm for LDS

- Initialize model parameters
  - $\Theta^{(0)} = \left\{\boldsymbol{A}^{(0)}, \boldsymbol{B}^{(0)}, \boldsymbol{C}^{(0)}, \boldsymbol{Q}^{(0)}, \boldsymbol{R}^{(0)}, \boldsymbol{m}_0^{(0)}, \boldsymbol{V}_0^{(0)}\right\}$
- Repeat until convergence:
  - [E-step] Inference by Kalman (RTS) Smoothing
    - Filtering : Compute $p\left(\boldsymbol{x}_t \middle| \boldsymbol{y}_{1:t}, \boldsymbol{u}_{1:t}, \Theta^{(t)}\right) = N(\boldsymbol{x}_t | \boldsymbol{m}_t, \boldsymbol{V}_t)$
    - Smoothing: Compute $p\left(\boldsymbol{x}_t \middle| \boldsymbol{y}_{1:T}, \boldsymbol{u}_{1:T}, \Theta^{(t)}\right) = N\left(\boldsymbol{x}_t \middle| \widehat{\boldsymbol{m}}_t, \widehat{\boldsymbol{V}}_t\right)$
    - Cross-covariance of $\boldsymbol{x}_t$ and $\boldsymbol{x}_{t+1}$ is also necessary
  - [M-step] Model update
    - $\Theta^{(t+1)} \leftarrow \underset{\Theta}{\operatorname{argmax}} E_{\boldsymbol{x}_{1:T}}[\log p(\boldsymbol{x}_{1:T}, \boldsymbol{y}_{1:T} | \boldsymbol{u}_{1:t}, \Theta)]$

Pros:

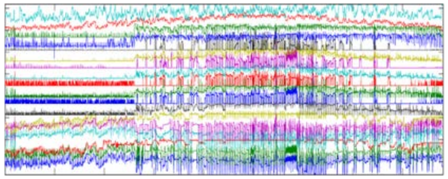
- Exact (for LDS), highly applicable (to complicated models)

Cons:

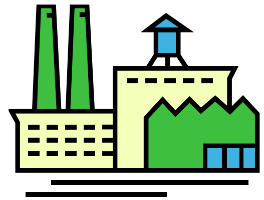
- Prone to local optima, many iterations (slow)

What is "subspace" identification method ?

Data



System



State Space Model

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{v}$$
$$\boldsymbol{y}_t = \boldsymbol{C}\boldsymbol{x}_t + \boldsymbol{w}$$

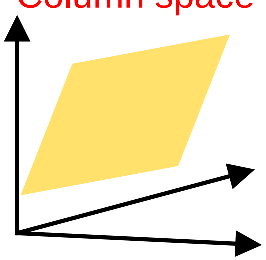
block Hankel matrix

$$\boldsymbol{Y}_f = \begin{bmatrix} \boldsymbol{y}_k & \boldsymbol{y}_{k+1} & \cdots & \boldsymbol{y}_{k+N-1} \\ \boldsymbol{y}_{k+1} & \boldsymbol{y}_{k+2} & \cdots & \boldsymbol{y}_{k+N} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{y}_{k+f-1} & \boldsymbol{y}_{k+f} & \cdots & \boldsymbol{y}_{k+f-2} \end{bmatrix}$$

extended observability matrix

$$\boldsymbol{\Gamma}_f = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \vdots \\ \boldsymbol{CA}^{f-1} \end{bmatrix}$$

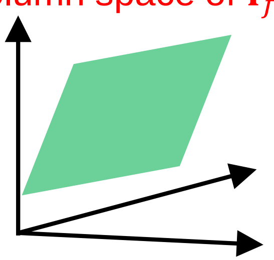
Column space of $\boldsymbol{Y}_f$



Equal under certain conditions

Identify

Column space of $\boldsymbol{\Gamma}_f$



- Advantages:
  - Global optimum is obtained without iteration by linear algebraic operations

Least Squares Linear Regression, and Singular Value Decomposition

Linear Regression

- Linear regression model (with vector output)

$$\boldsymbol{y} = \boldsymbol{\Gamma}\boldsymbol{x} + \boldsymbol{v}$$

error (residual)

Output (target) variable (vectors)

Input (exploratory) variable

- Collected data

$$\underbrace{[\boldsymbol{y}_1 \quad \boldsymbol{y}_2 \quad \cdots \quad \boldsymbol{y}_N]}_{\boldsymbol{Y}} = \boldsymbol{\Gamma}\underbrace{[\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_N]}_{\boldsymbol{X}} + \boldsymbol{V}$$

$$\boldsymbol{Y} = \color{red}{\boldsymbol{\Gamma}}\color{black}\boldsymbol{X} + \boldsymbol{V}$$

Estimate

Least Squares Solution

- Minimize sum of squares of residuals:

$$Loss = \|\boldsymbol{V}\|_{\mathrm{F}}^2 = \|\boldsymbol{Y} - \boldsymbol{\Gamma}\boldsymbol{X}\|_{\mathrm{F}}^2$$

$\|\cdot\|_{\mathrm{F}}^2$: Frobenius norm

$$= Tr[(\boldsymbol{Y} - \boldsymbol{\Gamma}\boldsymbol{X})^T(\boldsymbol{Y} - \boldsymbol{\Gamma}\boldsymbol{X})]$$

$$\frac{\partial}{\partial \boldsymbol{\Gamma}} Loss = -2\boldsymbol{Y}\boldsymbol{X}^T + \mathbf{2}\boldsymbol{\Gamma}\boldsymbol{X}\boldsymbol{X}^T = \mathbf{0}$$

Gradient is zero at the minimum

$$\hat{\boldsymbol{\Gamma}} = \boldsymbol{Y}\underbrace{\boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{X}^T\right)^{-1}}_{\text{Pseudo inverse of } \boldsymbol{X}}$$

Least square solution

$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{\Gamma}}\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}$$

Prediction

$$\boldsymbol{V} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y}\left(\boldsymbol{I} - \boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}\right)$$

Residual

Geometric Interpretation of Least Squares

Prediction

$$\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{\Gamma}}\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}$$

$$\equiv \boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{X}}$$

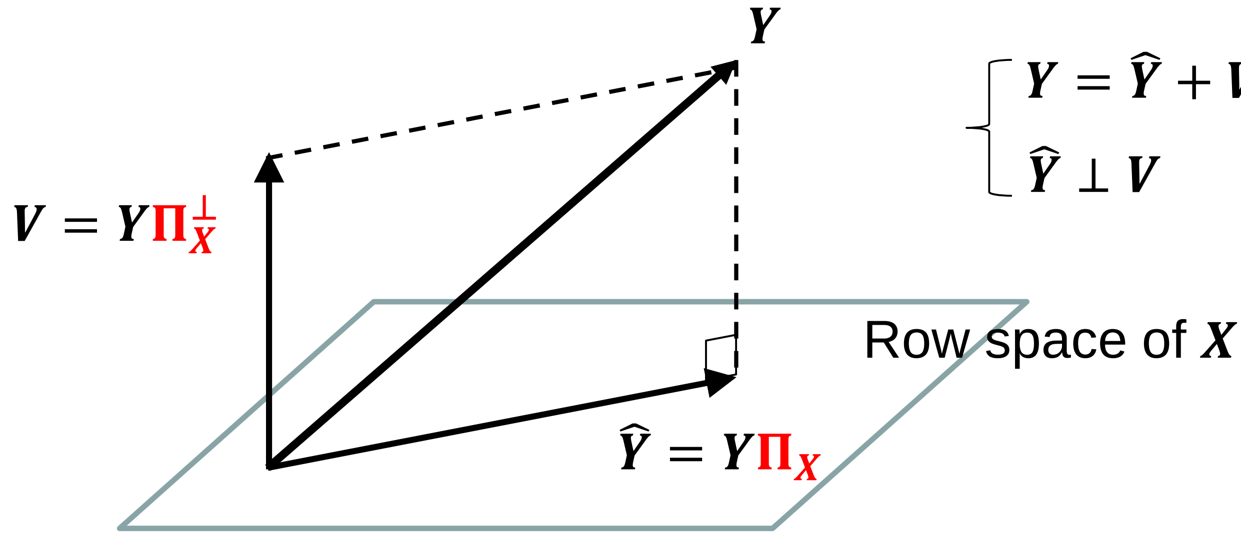
Orthogonal projection of $\boldsymbol{Y}$ onto the space spanned by the rows of $\boldsymbol{X}$ (row space)

Redidual

$$\boldsymbol{V} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}} = \boldsymbol{Y}\left(\boldsymbol{I} - \boldsymbol{X}^T\left(\boldsymbol{X}\boldsymbol{X}^T\right)^{-1}\boldsymbol{X}\right)$$

$$= \boldsymbol{Y}(\boldsymbol{I} - \boldsymbol{\Pi}_{\boldsymbol{X}}) \equiv \boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{X}}^{\perp}$$

Orthogonal projection of $\boldsymbol{Y}$ onto the complement of row space of $\boldsymbol{X}$



$$\begin{cases} \boldsymbol{Y} = \widehat{\boldsymbol{Y}} + \boldsymbol{V} \\ \widehat{\boldsymbol{Y}} \perp \boldsymbol{V} \end{cases}$$

When Two Sets of Explanatory Variables are given

- Consider there are two sets of inputs $\boldsymbol{X}$ and $\boldsymbol{U}$:

Unknown

$$\boldsymbol{Y} = \boldsymbol{\Gamma}\boldsymbol{X} + \boldsymbol{H}\boldsymbol{U} + \boldsymbol{V}$$

Known (data)

Residual

- Assume $\boldsymbol{V}$ is independent of both $\boldsymbol{X}$ and $\boldsymbol{U}$
- Multiply $\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp}$ from the right side

$$\boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp} = \boldsymbol{\Gamma}\boldsymbol{X}\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp} + \underbrace{\boldsymbol{H}\boldsymbol{U}\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp}}_{\mathbf{0}} + \underbrace{\boldsymbol{V}\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp}}_{\boldsymbol{V}}$$

$$\Rightarrow \boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp} = \boldsymbol{\Gamma}\boldsymbol{X}\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp} + \boldsymbol{V}$$

$$\Rightarrow \hat{\boldsymbol{\Gamma}} = \boldsymbol{Y}\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp}\boldsymbol{X}^{T}\left(\boldsymbol{X}\boldsymbol{\Pi}_{\boldsymbol{U}}^{\perp}\boldsymbol{X}^{T}\right)^{-1}$$

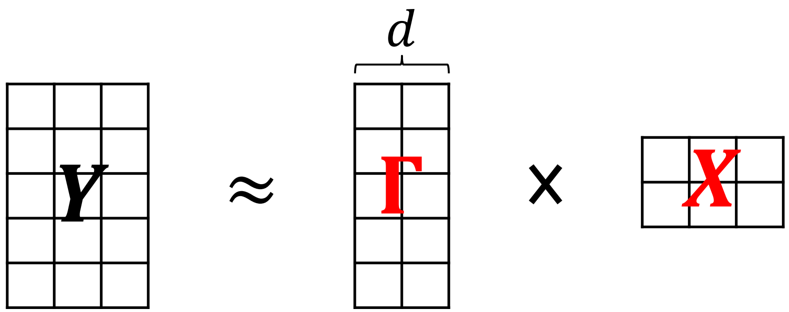
Remarks:
- $\hat{\boldsymbol{H}}$ can also be similarly obtained
- Known as oblique projection

What if $\boldsymbol{X}$ is not given ?

Consider both $\boldsymbol{\Gamma}$ and $\boldsymbol{X}$ are unknown

$$\boldsymbol{Y} = \boldsymbol{\Gamma}\boldsymbol{X} + \boldsymbol{W}$$

But assume the number of columns of $\boldsymbol{\Gamma}$ is known



This is the problem of low-rank approximation of a matrix

(Truncated) Singular Value Decomposition (SVD)

$$\boldsymbol{Y} = \boldsymbol{U}_d \boldsymbol{S}_d \boldsymbol{V}_d^T + \boldsymbol{W}$$

where, $\boldsymbol{S}_d = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_d \end{bmatrix}$

$$\begin{cases} \hat{\boldsymbol{\Gamma}} = \boldsymbol{U}_d \boldsymbol{S}_d^{1/2} \\ \hat{\boldsymbol{X}} = \boldsymbol{S}_d^{1/2} \boldsymbol{V}_d^T \end{cases}$$

(Note that solution is not unique)

Innovation Form of State Space Model

- Get back to the state space model

$$\begin{cases} \boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{v}_k \\ \boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{w}_k \end{cases}$$

- Kalman filter representation

$$\widehat{\boldsymbol{x}}_{k+1} = \boldsymbol{A}\widehat{\boldsymbol{x}}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{K}(\boldsymbol{y}_k - \boldsymbol{C}\widehat{\boldsymbol{x}}_k)$$

- Define the *innovation* : $\boldsymbol{e}_k = \boldsymbol{y}_k - \boldsymbol{C}\widehat{\boldsymbol{x}}_k$

$$\begin{cases} \widehat{\boldsymbol{x}}_{k+1} = \boldsymbol{A}\widehat{\boldsymbol{x}}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{K}\boldsymbol{e}_k \\ \boldsymbol{y}_k = \boldsymbol{C}\widehat{\boldsymbol{x}}_k + \boldsymbol{e}_k \end{cases}$$

Note that inputs $\boldsymbol{u}_k$ and outputs $\boldsymbol{y}_k$ are unchanged

Future Outputs

Innovation form

$$\begin{cases} \hat{\boldsymbol{x}}_{k+1} = \boldsymbol{A}\hat{\boldsymbol{x}}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{K}\boldsymbol{e}_k \\ \boldsymbol{y}_k = \boldsymbol{C}\hat{\boldsymbol{x}}_k + \boldsymbol{e}_k \end{cases}$$

1-step ahead $$\boldsymbol{y}_{k+1} = \boldsymbol{C}\hat{\boldsymbol{x}}_{k+1} + \boldsymbol{e}_{k+1} = \boldsymbol{C}\boldsymbol{A}\hat{\boldsymbol{x}}_k + \boldsymbol{C}\boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{C}\boldsymbol{K}\boldsymbol{e}_k + \boldsymbol{e}_{k+1}$$

⋮ $\quad\vdots$

j-step ahead $$\boldsymbol{y}_{k+j} = \boldsymbol{C}\boldsymbol{A}^j\hat{\boldsymbol{x}}_k + \begin{bmatrix}\boldsymbol{C}\boldsymbol{A}^{j-1}\boldsymbol{B} & \cdots & \boldsymbol{C}\boldsymbol{B} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{u}_k \\ \vdots \\ \boldsymbol{u}_{k+j}\end{bmatrix}$$

$$+\begin{bmatrix}\boldsymbol{C}\boldsymbol{A}^{j-1}\boldsymbol{K} & \cdots & \boldsymbol{C}\boldsymbol{K} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{e}_k \\ \vdots \\ \boldsymbol{e}_{k+j}\end{bmatrix}$$

stack vertically

Extended observability matrix

$$\begin{bmatrix} \boldsymbol{y}_k \\ \boldsymbol{y}_{k+1} \\ \vdots \\ \boldsymbol{y}_{k+f-1} \end{bmatrix} = \overbrace{\begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \vdots \\ \boldsymbol{CA}^{f-1} \end{bmatrix}}^{\boldsymbol{\Gamma}_f} \hat{\boldsymbol{x}}_k + \overbrace{\begin{bmatrix} \boldsymbol{0} & & & \\ \boldsymbol{CB} & \boldsymbol{0} & & \\ \vdots & \boldsymbol{CB} & \ddots & \\ \boldsymbol{CA}^{f-2}\boldsymbol{B} & \cdots & \boldsymbol{CB} & \boldsymbol{0} \end{bmatrix}}^{\boldsymbol{H}_f} \begin{bmatrix} \boldsymbol{u}_k \\ \boldsymbol{u}_{k+1} \\ \vdots \\ \boldsymbol{u}_{k+f-1} \end{bmatrix}$$

$$+ \underbrace{\begin{bmatrix} \boldsymbol{I} & & & \\ \boldsymbol{CK} & \boldsymbol{I} & & \\ \vdots & \boldsymbol{CK} & \ddots & \\ \boldsymbol{CA}^{f-2}\boldsymbol{K} & \cdots & \boldsymbol{CK} & \boldsymbol{I} \end{bmatrix}}_{\boldsymbol{G}_f} \begin{bmatrix} \boldsymbol{e}_k \\ \boldsymbol{e}_{k+1} \\ \vdots \\ \boldsymbol{e}_{k+f-1} \end{bmatrix}$$

Relation between the current state (estimate) $\hat{\boldsymbol{x}}_k$ and future measurements $\boldsymbol{y}_{k:k+f-1}$, given future inputs $\boldsymbol{u}_{k:k+f-1}$

$\boldsymbol{H}_f$ and $\boldsymbol{G}_f$ are Toeplitz matrices stack horizontally

Future output data matrix $\boldsymbol{Y}_f$

Kalman state sequence $\boldsymbol{X}_k$

Future input data matrix $\boldsymbol{U}_f$

$$\overbrace{\begin{bmatrix} \boldsymbol{y}_k & \cdots & \boldsymbol{y}_{k+N-1} \\ \vdots & \cdots & \vdots \\ \boldsymbol{y}_{k+f-1} & \cdots & \boldsymbol{y}_{k+f-2} \end{bmatrix}}^{\boldsymbol{Y}_f} = \boldsymbol{\Gamma}_f \overbrace{[\hat{\boldsymbol{x}}_k \quad \cdots \quad \hat{\boldsymbol{x}}_{k+N-1}]}^{\boldsymbol{X}_k} + \boldsymbol{H}_f \overbrace{\begin{bmatrix} \boldsymbol{u}_k & \cdots & \boldsymbol{u}_{k+N-1} \\ \vdots & \cdots & \vdots \\ \boldsymbol{u}_{k+f-1} & \cdots & \boldsymbol{u}_{k+f-2} \end{bmatrix}}^{\boldsymbol{U}_f}$$

$$+\boldsymbol{G}_f \underbrace{\begin{bmatrix} \boldsymbol{e}_k & \cdots & \boldsymbol{e}_{k+N-1} \\ \vdots & \cdots & \vdots \\ \boldsymbol{e}_{k+f-1} & \cdots & \boldsymbol{e}_{k+f-2} \end{bmatrix}}_{\boldsymbol{E}_f}$$

Future noise data matrix

Given | Given | Weighted residuals

$$\boldsymbol{Y}_f = \underbrace{\boldsymbol{\Gamma}_f \boldsymbol{X}_k}_{\text{Unknown}} + \boldsymbol{H}_f \boldsymbol{U}_f + \overbrace{\boldsymbol{G}_f \boldsymbol{E}_f}^{\text{Weighted residuals}}$$

$\boldsymbol{Y}_f$, $\boldsymbol{U}_f$, $\boldsymbol{E}_f$ are block Hankel matrices

Past and Future Data Matrices

Past input / output data matrices are also defined

	Output	Input
Past	$\overbrace{\begin{bmatrix} \boldsymbol{y}_{k-p} & \cdots & \boldsymbol{y}_{k-p+N-1} \\ \vdots & \cdots & \vdots \\ \boldsymbol{y}_{k-1} & \cdots & \boldsymbol{y}_{k+N-2} \end{bmatrix}}^{\boldsymbol{Y}_p}$	$\overbrace{\begin{bmatrix} \boldsymbol{u}_{k-p} & \cdots & \boldsymbol{u}_{k-p+N-1} \\ \vdots & \cdots & \vdots \\ \boldsymbol{u}_{k-1} & \cdots & \boldsymbol{u}_{k+N-2} \end{bmatrix}}^{\boldsymbol{U}_p}$
Future	$\underbrace{\begin{bmatrix} \boldsymbol{y}_{k} & \cdots & \boldsymbol{y}_{k+N-1} \\ \vdots & \cdots & \vdots \\ \boldsymbol{y}_{k+f-1} & \cdots & \boldsymbol{y}_{k+f-2} \end{bmatrix}}_{\boldsymbol{Y}_f}$	$\underbrace{\begin{bmatrix} \boldsymbol{u}_{k} & \cdots & \boldsymbol{u}_{k+N-1} \\ \vdots & \cdots & \vdots \\ \boldsymbol{u}_{k+f-1} & \cdots & \boldsymbol{u}_{k+f-2} \end{bmatrix}}_{\boldsymbol{U}_f}$

Kalman State Sequence

- Kalman state (sequence) $\boldsymbol{X}_k$ is unknown
- But, it is known it can be estimated from past inputs and outputs !

$$\boldsymbol{X}_k = \boldsymbol{L}_u \boldsymbol{U}_p + \boldsymbol{L}_y \boldsymbol{Y}_p = [\boldsymbol{L}_u \quad \boldsymbol{L}_y] \begin{bmatrix} \boldsymbol{U}_p \\ \boldsymbol{Y}_p \end{bmatrix} \equiv \boldsymbol{L}_z \boldsymbol{Z}_p$$

Past inputs and outputs

$$\boldsymbol{Y}_f = \boldsymbol{\Gamma}_f \boldsymbol{X}_k + \boldsymbol{H}_f \boldsymbol{U}_f + \boldsymbol{G}_f \boldsymbol{E}_f$$

$$\Downarrow$$

$$\boldsymbol{Y}_f = \boldsymbol{\Gamma}_f \boldsymbol{L}_z \boldsymbol{Z}_p + \boldsymbol{H}_f \boldsymbol{U}_f + \boldsymbol{G}_f \boldsymbol{E}_f$$

Open Loop Assumption

- We assume the system is open loop
- Future innovation is uncorrelated to future inputs

$$\boldsymbol{E}_f \boldsymbol{U}_f{}^T = \boldsymbol{0} \quad \Leftrightarrow \quad \boldsymbol{E}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} = \boldsymbol{E}_f$$

- Future innovation is uncorrelated to past inputs and outputs

$$\boldsymbol{E}_f \boldsymbol{Z}_p{}^T = \boldsymbol{0}$$

Subspace Identification 1 (N4SID approach) [Qin 2004][Qin 2006]

- Data matrices $\boldsymbol{Y}_f$, $\boldsymbol{U}_f$ and $\boldsymbol{Z}_p$ are given

$$\boldsymbol{Y}_f = \boldsymbol{\Gamma}_f \boldsymbol{L}_z \boldsymbol{Z}_p + \boldsymbol{H}_f \boldsymbol{U}_f + \boldsymbol{G}_f \boldsymbol{E}_f$$

- Multiply $\boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp}$ from the right side

$$\boldsymbol{Y}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} = \boldsymbol{\Gamma}_f \boldsymbol{L}_z \boldsymbol{Z}_p \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} + \boldsymbol{H}_f \overbrace{\boldsymbol{U}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp}}^{\text{Zero}} + \boldsymbol{G}_f \overbrace{\boldsymbol{E}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp}}^{\boldsymbol{E}_f}$$

$$= \boldsymbol{\Gamma}_f \boldsymbol{L}_z \boldsymbol{Z}_p \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} + \boldsymbol{G}_f \boldsymbol{E}_f$$

- Multiply $\boldsymbol{Z}_p{}^T$ from the right side

$$\underbrace{\boldsymbol{Y}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} \boldsymbol{Z}_p{}^T}_{\text{Known (from data)}} = \boldsymbol{\Gamma}_f \boldsymbol{L}_z \boldsymbol{Z}_p \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} \boldsymbol{Z}_p{}^T + \boldsymbol{G}_f \overbrace{\boldsymbol{E}_f \boldsymbol{Z}_p{}^T}^{\text{Zero}}$$

$$= \boldsymbol{\Gamma}_f \boldsymbol{L}_z \boldsymbol{Z}_p \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} \boldsymbol{Z}_p{}^T$$

N4SID Approach (Continued)

- Perform (truncated) SVD on $\boldsymbol{Y}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} {\boldsymbol{Z}_p}^T$

$$\boldsymbol{Y}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} {\boldsymbol{Z}_p}^T \approx \boldsymbol{U}_d \boldsymbol{S}_d \boldsymbol{V}_d^T \left(\approx \boldsymbol{\Gamma}_f \boldsymbol{L}_z \boldsymbol{Z}_p \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} {\boldsymbol{Z}_p}^T \right)$$

- Determine the extended observability matrix as:

$$\hat{\boldsymbol{\Gamma}}_f \leftarrow \boldsymbol{U}_d {\boldsymbol{S}_d}^{1/2}$$

- Remember $\boldsymbol{\Gamma}_f = \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{C}\boldsymbol{A} \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}^{f-1} \end{bmatrix}$

⇨ System matrices $\boldsymbol{A}$, $\boldsymbol{C}$ are obtained

Subspace Identification 2 : Regression Approach

- Multiply $\boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f}$ to $\boldsymbol{Y}_f = \boldsymbol{\Gamma}_f \boldsymbol{L}_z \boldsymbol{Z}_p + \boldsymbol{H}_f \boldsymbol{U}_f + \boldsymbol{G}_f \boldsymbol{E}_f$

$$\underbrace{\boldsymbol{Y}_f \boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f}}_{\text{Known}} = \underbrace{\boldsymbol{\Gamma}_f \boldsymbol{L}_z}_{\text{Unknown}} \underbrace{\boldsymbol{Z}_p \boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f}}_{\text{Known}} + \underbrace{\boldsymbol{G}_f \boldsymbol{E}_f}_{\text{Residual}}$$

Same as the first approach

Smaller is better

Target variables

Explanatory variables

Obtained by (multi-variate) linear regression:

$$\widehat{\boldsymbol{\Gamma}_f \boldsymbol{L}_z} = \boldsymbol{Y}_f \boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f} \left(\boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f} \boldsymbol{Z}_p{}^T\right) \left(\boldsymbol{Z}_p \boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f} \boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f} \boldsymbol{Z}_p{}^T\right)^{-1}$$

$$= \boldsymbol{Y}_f \boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f} \boldsymbol{Z}_p{}^T \left(\boldsymbol{Z}_p \boldsymbol{\Pi}^{\perp}_{\boldsymbol{U}_f} \boldsymbol{Z}_p{}^T\right)^{-1} \approx \boldsymbol{U}_d \boldsymbol{S}_d \boldsymbol{V}_d^T \quad \text{(SVD)}$$

$$\hat{\boldsymbol{\Gamma}}_f \leftarrow \boldsymbol{U}_d \boldsymbol{S}_d{}^{1/2}$$

Subspace Identification 3 : CCA Approach*

$$\underbrace{\boldsymbol{Y}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp}}_{\text{Known}} = \underbrace{\boldsymbol{\Gamma}_f \boldsymbol{L}_z}_{\text{Unknown}} \underbrace{\boldsymbol{Z}_p \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp}}_{\text{Known}} + \underbrace{\boldsymbol{G}_f \boldsymbol{E}_f}_{\text{Residual}}$$

Variable set 1 Variable set 2

Obtained by Canonical Correlation Analysis (CCA):

Define $\boldsymbol{W}_r = \left(\boldsymbol{Y}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} \boldsymbol{Y}_f{}^{\boldsymbol{T}}\right)^{-1/2}$ and $\boldsymbol{W}_c = \left(\boldsymbol{Z}_p \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} \boldsymbol{Z}_p{}^{\boldsymbol{T}}\right)^{-1/2}$

Perform SVD on $\boldsymbol{W}_r \boldsymbol{Y}_f \boldsymbol{\Pi}_{\boldsymbol{U}_f}^{\perp} \boldsymbol{Z}_p{}^{\boldsymbol{T}} \boldsymbol{W}_c \approx \boldsymbol{U}_d \boldsymbol{S}_d \boldsymbol{V}_d^T$

Then, $\hat{\boldsymbol{\Gamma}}_f \leftarrow \boldsymbol{W}_r{}^{-1} \boldsymbol{U}_d \boldsymbol{S}_d{}^{1/2}$

(*) Also known as canonical variate analysis (CVA)

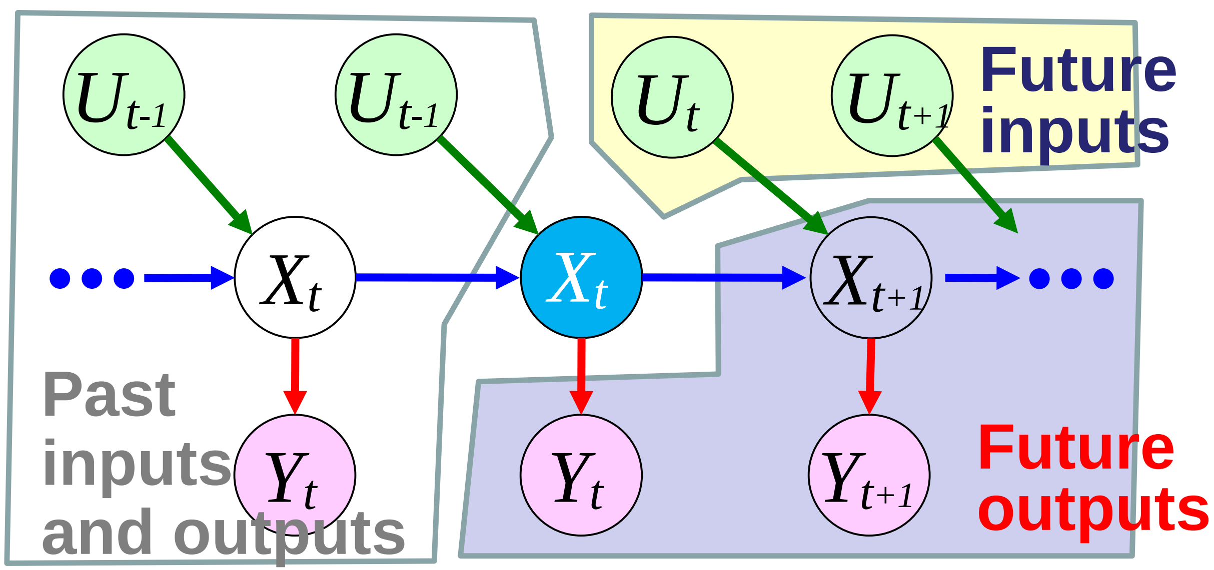
CCA-based Subspace Identification (Continued)

- Additionally, state vectors are obtained by

$$\hat{\boldsymbol{X}}_k = \boldsymbol{S}_d{}^{1/2} \boldsymbol{V}_d^T \boldsymbol{W}_c \boldsymbol{Z}_p$$

- This result is very interesting, because the "state" is obtained by CCA between "past" and "future" data sets.

- This idea goes back to [Akaike 1975].



"The state vector carries information necessary to predict the future output based on the past." [Katayama 2005]

Reference

- [Qin 2006] S. Joe Qin, "An overview of subspace identification", Computers & Chemical Engineering, Volume 30, Issues 10–12, 2006, Pages 1502-1513
  - Presentation file [Qin 2004] is also available (http://people.duke.edu/~hpgavin/SystemID/References/Qin-SubspaceID-2004.pdf)
- [Overschee 1996] Peter VAN OVERSCHEE & Bart DE MOOR, "SUBSPACE IDENTIFICATION FOR LINEAR SYSTEMS - Theory - Implementation - Applications", Springer, 1996
- [Katayama 2005] Toru Katayama, "Subspace Methods for System Identification ", Springer, 2005
- [Akaike 1975] H.Akaike, "Markovian Representation of Stochastic Processes by Canonical Variables", SIAM J. CONTROL, Vol. 13, No. 1, pp.162-173, January 1975

Appendix:
Canonical Correlation Analysis (CCA)

Canonical Correlation Analysis (1)

There are two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$

$\mathbf{x} \in R^m$ (m-dim.) $\mathbf{y} \in R^n$ (n-dim.)

For simplicity, both $\boldsymbol{x}$ and $\boldsymbol{y}$ are "centered", i.e., their means are zero vectors.

$$\mathrm{E}\left[\mathbf{x}\right] = \mathbf{0}_m \qquad \mathrm{E}\left[\mathbf{y}\right] = \mathbf{0}_n$$

Define covariance matrices of / between $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\mathrm{var}(\mathbf{x}) = \mathrm{E}\left[\mathbf{x}\mathbf{x}^T\right] = \mathbf{\Sigma}_{xx} \qquad \mathrm{var}(\mathbf{y}) = \mathrm{E}\left[\mathbf{y}\mathbf{y}^T\right] = \mathbf{\Sigma}_{yy}$$

$$\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \mathrm{E}\left[\mathbf{x}\mathbf{y}^T\right] = \mathbf{\Sigma}_{xy}$$

Canonical Correlation Analysis (2)

Consider to construct synthetic variable u and v by linear combinations of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively.

$$u = \mathbf{a}^T \mathbf{x} = a_1 x_1 + a_2 x_2 + \cdots + a_m x_m$$
$$v = \mathbf{b}^T \mathbf{y} = b_1 y_1 + b_2 y_2 + \cdots + b_n y_n$$

Problem: Find $\boldsymbol{a}$ and $\boldsymbol{b}$ so that the correlation between u and v is maximized

$$\rho = \mathbf{cor}(u, v) = \frac{\mathrm{cov}(u, v)}{\sqrt{\mathrm{var}(u)}\sqrt{\mathrm{var}(v)}}$$

$$= \frac{\mathbf{a}^T \boldsymbol{\Sigma}_{xy} \mathbf{b}}{\sqrt{\mathbf{a}^T \boldsymbol{\Sigma}_{xx} \mathbf{a}}\sqrt{\mathbf{b}^T \boldsymbol{\Sigma}_{yy} \mathbf{b}}}$$

Canonical Correlation Analysis (3)

Impose constraints on $\boldsymbol{a}$ and $\boldsymbol{b}$, so that the variances of u and v become 1

$$\operatorname{var}(u) = \operatorname{var}\left(\mathbf{a}^T \mathbf{x}\right) = \mathbf{a}^T \mathbf{\Sigma}_{xx} \mathbf{a} = 1$$

$$\operatorname{var}(v) = \operatorname{var}\left(\mathbf{b}^T \mathbf{y}\right) = \mathbf{b}^T \mathbf{\Sigma}_{yy} \mathbf{b} = 1$$

The goal of CCA is to solve:

$$(\boldsymbol{a}_1, \boldsymbol{b}_1) = \underset{\boldsymbol{a}^T \boldsymbol{\Sigma}_{xx} \boldsymbol{a} = 1, \boldsymbol{b}^T \boldsymbol{\Sigma}_{yy} \boldsymbol{b} = 1}{\operatorname{argmax}} \boldsymbol{a}^T \boldsymbol{\Sigma}_{xy} \boldsymbol{b}$$

It can be solved by Lagrange multiplier, but there is a more elegant method.

Canonical Correlation Analysis (4)

Let square root matrices of Σ_{xx} and Σ_{yy} be ${\Sigma_{xx}}^{1/2}$ and ${\Sigma_{yy}}^{1/2}$, respectively. Then, define $\boldsymbol{c}$ and $\boldsymbol{d}$ as,

$$\boldsymbol{c} = {\boldsymbol{\Sigma}_{xx}}^{1/2}\boldsymbol{a} \quad \text{, and } \boldsymbol{d} = {\boldsymbol{\Sigma}_{yy}}^{1/2}\boldsymbol{b}$$

The problem turns to be

$$(\boldsymbol{c}_1, \boldsymbol{d}_1) = \underset{\|\boldsymbol{c}\|=1\ ,\|\boldsymbol{d}\|=1}{\operatorname{argmax}} \boldsymbol{c}^T \left(\boldsymbol{\Sigma}_{xx}^{-T/2}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{xx}^{-1/2}\right)\boldsymbol{d}$$

This problem can be solved by SVD !

$$\boldsymbol{\Sigma}_{xx}^{-T/2}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{xx}^{-1/2} = \boldsymbol{USV}^T \quad \Rightarrow \quad \boldsymbol{c}_1 = \boldsymbol{u}_1, \boldsymbol{d}_1 = \boldsymbol{v}_1$$

$$\Rightarrow \quad \boldsymbol{a}_1 = {\boldsymbol{\Sigma}_{xx}}^{1/2}\boldsymbol{u}_1, \text{ and } \boldsymbol{b}_1 = {\boldsymbol{\Sigma}_{yy}}^{1/2}\boldsymbol{v}_1$$