System Estimation Methods III: Subspace Identification

Advanced Topics

Today we will see a more detailed derivation of the subspace methods

REMINDER: PRELIMINARIES

System (MIMO):

$$x_{t+1} = Ax_t + Bu_t + w_t$$
$$y_t = Cx_t + Du_t + v_t$$

Input:	$u_t \in \mathbb{R}^m$	Process noise:	$w_t \in \mathbb{R}^n$
State:	$x_t \in \mathbb{R}^n$	Measurement noise:	$v_t \in \mathbb{R}^p$
Output:	$y_t \in \mathbb{R}^p$		

Assumptions: $\{w_t\}$ and $\{v_t\}$ are white noise sequences

There are no constraints on A, B, C, D

REMINDER: PREDICTORS

$$\boxed{Y_t^r = O^r x_t + S^r U_t^r + V_t} \qquad \textbf{Fundamental Equation}$$

where

$$Y^r := \begin{bmatrix} y_t \\ y_{t+1} \\ \vdots \\ y_{t+r-1} \end{bmatrix}, \quad U^r := \begin{bmatrix} u_t \\ u_{t+1} \\ \vdots \\ u_{t+r-1} \end{bmatrix}, \quad V_t := \begin{bmatrix} v_t \\ Cw_t + v_{t+1} \\ \vdots \\ CA^{r-2}w_t + CA^{r-3}w_{t+1} + \cdots + Cw_{t+r-2} + v_{t+r-1} \end{bmatrix}$$

$$O^r := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix}, \qquad S^r := \begin{bmatrix} D & 0 & \cdots & 0 & 0 \\ CB & D & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ CA^{r-2}B & CA^{r-3}B & \cdots & CB & D \end{bmatrix}$$

ESTIMATION OF THE (EXTENDED) OBSERVABILITY MATRIX

This equation can be written as

$$\boxed{\mathbf{Y} = O^r \mathbf{X} + S^r \mathbf{U} + \mathbf{V}}$$

where

$$\begin{aligned}
\mathbf{Y} &:= [Y_1^r \quad \cdots \quad Y_N^r] \\
\mathbf{X} &:= [x_1 \quad \cdots \quad x_N] \\
U &:= [U_1^r \quad \cdots \quad U_N^r] \\
\mathbf{V} &:= [V_1 \quad \cdots \quad V_N]
\end{aligned}$$

Objective: To estimate $O^r \mathbf{X}$, given data $\mathbf{U}$ and $\mathbf{Y}$.

Remark: A state transformation $A \to T^{-1}AT$, $C \to CT$ changes O^r to $O^r T$.
Thus, postmultiplying O^r by an invertible T simply changes the resulting realization

ESTIMATION OF THE (EXTENDED) OBSERVABILITY MATRIX (CONT.)

To eliminate $\mathbf{U}$, post-multiply by a *projector* $\Pi_{\mathbf{U}^\perp}^\perp := I - \mathbf{U}^T(\mathbf{U}\mathbf{U}^T)^{-1}\mathbf{U}$, giving

$$\mathbf{Y}\Pi_{\mathbf{U}^\perp}^\perp = O^r\mathbf{X}\Pi_{\mathbf{U}^\perp}^\perp + S^r\mathbf{U}\Pi_{\mathbf{U}^\perp}^\perp + \mathbf{V}\Pi_{\mathbf{U}^\perp}^\perp = O^r\mathbf{X}\Pi_{\mathbf{U}^\perp}^\perp + \mathbf{V}\Pi_{\mathbf{U}^\perp}^\perp$$

The noise term, $\mathbf{V}\Pi_{\mathbf{U}^\perp}^\perp$, can be eliminated using *instrumental variables*, i.e., post-multiplying by a matrix $\Phi^T := [\varphi_1^s \quad \cdots \quad \varphi_N^s]^T \in \mathbb{R}^{N\times s}$, so that

$$\boxed{\underbrace{\frac{1}{N}\mathbf{Y}\Pi_{\mathbf{U}^\perp}^\perp\Phi^T}_{G} = O^r\underbrace{\frac{1}{N}\mathbf{X}\Pi_{\mathbf{U}^\perp}^\perp\Phi^T}_{\tilde{T}_N} + \underbrace{\frac{1}{N}\mathbf{V}\Pi_{\mathbf{U}^\perp}^\perp\Phi^T}_{V_N} \xrightarrow{N\to\infty} O^r\tilde{T}}$$

ESTIMATION OF THE (EXTENDED) OBSERVABILITY MATRIX (CONT.)

Therefore, we need

$$
\begin{aligned}
0 &= \lim_{N\to\infty} V_N \\
&= \lim_{N\to\infty} \frac{1}{N} \mathbf{V}\Pi^{\perp}_{\mathbf{U}^{\perp}} \Phi^T \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} V_t (\varphi_t^s)^T - \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} V_t (U_t^r)^T \left[\frac{1}{N} \sum_{t=1}^{N} U_t^r (U_t^r)^T \right]^{-1} \frac{1}{N} \sum_{t=1}^{N} U_t^r (\varphi_t^s)^T \\
&= \bar{E}\{V_t (\varphi_t^s)^T\} - \bar{E}\{V_t (U_t^r)^T\} \left[\bar{E}\{U_t^r (U_t^r)^T\} \right]^{-1} \bar{E}\{U_t^r (\varphi_t^s)^T\}
\end{aligned}
$$

In open loop, $\{U_t^r\}$ is independent of $\{V_t\}$, so $\bar{E}\{V_t (U_t^r)^T\} = 0$. The first term is 0 if we build φ_t^s from past data, e.g.,

$$
\varphi_t^s = [y_{t-1} \quad \cdots \quad y_{t-s_1} \quad u_{t-1} \quad \cdots \quad u_{t-s_2}]^T
$$

ESTIMATION OF THE (EXTENDED) OBSERVABILITY MATRIX (CONT.)

We also need $\tilde{T} = \lim_{N \to \infty} N^{-1} \mathbf{X} \Pi_{\mathbf{U}^{\perp}}^{\perp} \Phi^T$ nonsingular. For the previous choice of φ_t^s this holds under some conditions (see Problem 10G.6 of Ljung)

ESTIMATION OF THE ORDER

If we have a noisy estimation $G = O^r T + E_N$, where E_N is small, then rank$\{O^r\}$ can be estimated via SVD:

$$G = USV^T = U \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{n^*} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} V^T$$

The smaller σ_i's ($< \varepsilon$ for some predetermined $\varepsilon > 0$) can be replaced by 0, thus replacing G by a lower rank matrix $G_1 = U_1 S_1 V_1^T$. From a previous remark, only U_1 is important for estimating A, B, C, D

ESTIMATION OF THE ORDER (CONT.)

Many methods using weighting matrices for the SVD step, i.e.,

$$W_1 G W_2 = USV^T \approx U_1 S_1 V_1^T$$

and then consider: $\hat{O}^r = W_1^{-1} U_1 R$

where R is an arbitrary matrix (to determine a particular state realization)
W_2 corresponds to a state transformation
W_1 only affects $\hat{O}^r$ when there is noise, so it affects the quality of $\hat{A}, \hat{C}$

Typical choices: $R = I$, $R = S_1$ or $R = S_1^{1/2}$

MOESP	$W_1 = I$, $W_2 = (N^{-1}\Phi\Pi_{U^T}^{\perp}\Phi^T)^{-1}\Phi\Pi_{U^T}^{\perp}$
N4SID	$W_1 = I$, $W_2 = (N^{-1}\Phi\Pi_{U^T}^{\perp}\Phi^T)^{-1}\Phi$
IVM	$W_1 = (N^{-1}Y\Pi_{U^T}^{\perp}Y)^{-1/2}$, $W_2 = (N^{-1}\Phi\Phi^T)^{-1/2}$
CVA	$W_1 = (N^{-1}Y\Pi_{U^T}^{\perp}Y)^{-1/2}$, $W_2 = (N^{-1}\Phi\Pi_{U^T}^{\perp}\Phi^T)^{-1/2}$

ESTIMATION OF A AND C

A and C can be estimated from the (extended) observability matrix

$$O^r := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix}$$

by solving the equations:

$$\boxed{\begin{aligned} \hat{C} &= O^r(1:p, 1:n) \\ O^r(p+1:pr, 1:n) &= O^r(1:p(r-1), 1:n)\hat{A} \end{aligned}}$$

ESTIMATION OF B AND D

Given $\hat{A}$ and $\hat{C}$, we can estimate B and D (and the initial state x_0) via LS from:

$$\boxed{y_t = \hat{C}(qI - \hat{A})^{-1} x_0 \delta_t + \hat{C}(qI - \hat{A})^{-1} B u_t + D u_t + \varepsilon_t}$$

where $\varepsilon_t := \hat{C}(qI - \hat{A})^{-1} w_t + v_t$.

Remark: It is possible to find the state x_t, and from this to estimate the noise statistics. For more details, see Ljung, pp. 348-349.

Summary of Subspace Methods

1. From data, form $G = N^{-1}\mathbf{Y}\Pi^{\perp}_{\mathbf{U}^T}\Phi^T$

2. Choose W_1, W_2 and perform SVD: $\hat{G} = W_1 G W_2 = USV^T \approx U_1 S_1 V_1^T$

3. Select R and define $\hat{O}^r = W_1^{-1} U_1 R$, from which estimate $\hat{A}$, $\hat{C}$ via

$$\hat{C} = O^r(1:p, 1:n)$$

$$O^r(p+1:pr, 1:n) = O^r(1:p(r-1), 1:n)\hat{A}$$

4. Estimate $\hat{B}$, $\hat{D}$ via LS from

$$y_t = \hat{C}(qI - \hat{A})^{-1} x_0 \delta_t + \hat{C}(qI - \hat{A})^{-1} B u_t + D u_t + \varepsilon_t$$