SYSTEM IDENTIFICATION

Ali Karimpour
Assistant Professor
Ferdowsi University of Mashhad

Reference: “System Identification Theory For The User”
Lennart Ljung
Lecture 12

State Space System Identification
Subspace Methods for Estimating State Space Models

Let us now consider how to estimate the system matrices $A$, $B$, $C$ and $D$ in the ss model

$$
x(t+1) = Ax(t) + Bu(t) + w(t)$$
$$y(t) = Cx(t) + Du(t) + v(t)$$

Let the output $y(t)$ is a $p$-dimensional column vector, the input $u(t)$ is a $m$-dimensional column vector. Also the order of system is $n$.

We also assume that this $ss$ representation is a minimal realization.

We know that many different representation can also described the system. They are:

$$\tilde{x}(t+1) = T^{-1} A T \tilde{x}(t) + T^{-1} B u(t) + \tilde{w}(t)$$
$$y(t) = C T \tilde{x}(t) + D u(t) + \tilde{v}(t)$$

Where $T$ is any invertible matrix. We also have

$$\tilde{x}(t) = T^{-1} x(t)$$
Subspace Methods for Estimating State Space Models

Let the ss as:

\[ x(t + 1) = Ax(t) + Bu(t) + w(t) \]
\[ y(t) = Cx(t) + Du(t) + v(t) \]

I) If \( \hat{A} \) and \( \hat{C} \) are known. It is easy to find \( B \) and \( D \).

- ► Estimating \( B \) and \( D \)  
  But \( \hat{A} \) and \( \hat{C} \) are not known.

II) If the (extended) observability matrix, \( O_r \), for the system is known.

  It is easy to find \( A \) and \( C \).

- ► Finding \( A \) and \( C \) from Observability matrix  
  But \( O_r \) is not known.

III) The extended observability matrix can be consistently estimated.

- ► Estimating the Extended Observability matrix

IV) Once the observability matrix has been estimated, the state can be constructed.

- ► Finding the States and Estimating the noise Statistics.
Subspace Methods for Estimating State Space Models

Let the ss as:

\[ x(t + 1) = Ax(t) + Bu(t) + w(t) \]
\[ y(t) = Cx(t) + Du(t) + v(t) \]

- Estimating \( B \) and \( D \)
  
  Suppose \( \hat{A} \) and \( \hat{C} \) are known → Try to find \( B \) and \( D \)

- Finding \( A \) and \( C \) from Observability matrix
  
  Suppose \( O_r \) is known → Try to find \( A \) and \( C \)

- Estimating the Extended Observability matrix
  
  Suppose input and outputs are available → Try to find \( O_r \)

- Finding the States and Estimating the noise Statistics.
Subspace Methods for Estimating State Space Models

▶ Estimating \( B \) and \( D \)

Suppose \( \hat{A} \) and \( \hat{C} \) are known → Try to find \( B \) and \( D \).

For given and fixed \( \hat{A} \) and \( \hat{C} \) the model structure:

\[
\begin{align*}
x(t + 1) &= \hat{A}x(t) + Bu(t) + w(t) \\
y(t) &= \hat{C}x(t) + Du(t) + v(t)
\end{align*}
\]

\[
y(t) = \hat{C}(qI - A)^{-1} Bu(t) + Du(t) + \frac{\hat{C}(qI - A)^{-1}w(t) + v(t)}{v(t)}
\]

\[
\hat{y}(t | B, D) = \hat{C}(qI - A)^{-1} Bu(t) + Du(t)
\]

It is clearly linear in \( B \) and \( D \).

If the system operates in open loop. We can thus consistently estimate \( B \) and \( D \) according to theorem 8.4 even if the noise sequence is non-white.
Theorem 8.4. Suppose that the data set $Z^\infty$ is subject to assumptions D1 and S1. Let $\mathcal{M}$ be a linear uniformly stable model structure, such that $G$ and $H$ are independently parametrized:

$$
\theta = \begin{bmatrix} \rho \\ \eta \end{bmatrix}, \quad G(q, \theta) = G(q, \rho), \quad H(q, \theta) = H(q, \eta) \quad (8.45)
$$

and such that the set

$$
D_G(S, \mathcal{M}) = \{ \rho | G(e^{i\omega}, \rho) = G_0(e^{i\omega}) \ \forall \omega \} \quad (8.46)
$$
is nonempty. Assume that $Z^\infty$ is informative enough with respect to $\mathcal{M}$ and that the system operates in open loop; that is,

$$
\{u(t)\} \text{ and } \{e_0(t)\} \text{ are independent} \quad (8.47)
$$

Let

$$
\hat{\theta}_N = \begin{bmatrix} \hat{\rho}_N \\ \hat{\eta}_N \end{bmatrix}
$$

be obtained by the prediction-error method (8.16) and (8.17). Then

$$
\hat{\rho}_N \to D_G(S, \mathcal{M}) \quad \text{w.p. 1 as } N \to \infty \quad (8.48)
$$

The result (8.48) can be written more suggestively as

$$
G(e^{i\omega}, \hat{\theta}_N) \to G_0(e^{i\omega}), \quad \text{w.p. 1 as } N \to \infty \quad (8.49)
$$
Subspace Methods for Estimating State Space Models

**Estimating B and D** Suppose \( \hat{A} \) and \( \hat{C} \) are known \( \rightarrow \) Try to find \( B \) and \( D \)

Let us write the predictor in the standard linear regression form

\[
x(t+1) = \hat{A}x(t) + Bu(t) = \hat{A}x(t) + \begin{bmatrix} b_1 & \ldots & b_m \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} = \hat{A}x(t) + b_1 u_1 + \ldots + b_m u_m
\]

\[
y(t) = \hat{C}x(t) + Du(t) = \hat{C}x(t) + \begin{bmatrix} d_1 & \ldots & d_m \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} = \hat{C}x(t) + d_1 u_1 + \ldots + d_m u_m
\]

\[
\hat{y}(t) = \hat{C}(sI - \hat{A})^{-1} \left( b_1 u_1 + \ldots + b_m u_m \right) + d_1 u_1 + \ldots + d_m u_m
\]

\[
\hat{y}(t) = \varphi(t)\theta = \varphi(t) \begin{bmatrix} \text{Vec}(B) \\ \text{Vec}(D) \end{bmatrix}
\]
Subspace Methods for Estimating State Space Models

**Estimating \( B \) and \( D \)**

Suppose \( \hat{A} \) and \( \hat{C} \) are known → Try to find \( R \) and \( D \)

\[
\hat{y}(t) = \hat{C}(sI - \hat{A})^{-1}(b_1 u_1 + ... + b_m u_m) + d_1 u_1 + ... + d_m u_m
\]

\[
\begin{bmatrix}
\hat{y}(t) \\
\varphi(t) \theta \\
\end{bmatrix} = \varphi(t) \begin{bmatrix}
\text{Vec}(B) \\
\text{Vec}(D)
\end{bmatrix}
\]

For \( p \times mn + mp \) input, \( mn + mp \) parameters

To find the \( r \):th (\( r \leq mn \)) column of \( \varphi(t) \), which corresponds to the \( r \):th element of \( \theta \), i.e., the element \( B_{j,k} \), we differentiate (10.85b) w.r.t. this element and obtain

\[
\varphi_r(t) = \hat{C}(qI - \hat{A})^{-1} E_j u_k(t)
\]

where \( E_j \) is the column vector with the \( j \):th element equal to 1 and the others equal to 0. The rows for \( r > nm \) are handled in a similar way.

Clearly \( B \) and \( D \) derived by simple LS method.
Subspace Methods for Estimating State Space Models

► Estimating $x_0$

If desired, also the initial state $x_0 = x(0)$ can be estimated in an analogous way, since the predictor with initial values taken into account is

$$
\hat{y}(t \mid B, D, x_0) = \hat{C}(qI - A)^{-1} x_0 \delta(t) + \hat{C}(qI - A)^{-1} Bu(t) + Du(t)
$$

Which is linear also in $x_0$. Here $\delta(t)$ is the unit pulse at time 0.

Remark: If $\hat{A}$ and $\hat{C}$ are the correct values, the least squares estimates of $B$ and $D$ will also converge to their true values, according to Theorem 8.4. If consistent estimates $\hat{A}_N$ and $\hat{C}_N$ are used instead, convergence of $\hat{B}_N$ and $\hat{D}_N$ to their true values still holds. This follows by fairly straightforward calculations. See Vandersteen, Van hamme, and Pintelon (1996) for a general treatment of such issues.
Subspace Methods for Estimating State Space Models

Finding $A$ and $C$ from Observability matrix

Suppose $O_r \rightarrow$ Try to find $A$ and $C$

Suppose $pr \times n^*$ $G$ that is:

\[ G(O_r) \]

There is two situation:

- Known System Order.
- Unknown System Order.

**Known System Order.** Suppose first we know that

\[ G = O_r \]

So that $n^* = n$. To find $C$ is then immediate:

\[ C = O_r (1:p,1:n) \]
Finding $A$ and $C$ from Observability matrix

Suppose $O_r \rightarrow$ Try to find $A$ and $C$

Similarly, we can find $\hat{A}$ from the equation

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

$O_r = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-2} \\ CA^{r-1} \end{bmatrix}$

Under the observability assumption, $O_{r-1}$ has rank $n$ so $\hat{A}$ can be determined uniquely.

$n^2$ unknown in $\hat{A}$

$np(r-1)$ equations
Subspace Methods for Estimating State Space Models

Finding $A$ and $C$ from Observability matrix

Suppose $O_r \rightarrow$ Try to find $A$ and $C$

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

$O_{up} \hat{A} = O_{down}$

$\hat{A} = \left( O_{up}^T O_{up} \right)^{-1} O_{up}^T O_{down}$

Role of the State Space Basis

The extended observability matrix is depends on the choice of basis in the state-space representation. It is easy to verify that the observability matrix would be

$\tilde{O}_r = O_r T$

So, multiplying the extended observability matrix from right, just changes the basis representation.
Unknown system order.

Suppose now the true orders of the system is unknown. And that \( n^* \)-the number of columns of \( G \) is just an upper bound for the order. 

\[
pr \times n^* \quad n \times n^*
\]

\[
G = O_r \tilde{T}
\]

where also \( n \) is unknown to us.

The rank of \( G \) is \( n \). A straightforward way is reduce the column of \( G \) by \( n \).

A general and numerically sound way of reducing the column space is to use singular value decomposition (SVD): 

\[
G = USV^T = U \begin{bmatrix}
\sigma_1 & 0 & 0 & \ldots & 0 \\
0 & \sigma_2 & 0 & \ldots & 0 \\
0 & 0 & \sigma_3 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \sigma_{n^*} \\
0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0
\end{bmatrix} V^T
\]
Subspace Methods for Estimating State Space Models

\[
G = USV^T = U \begin{bmatrix}
\sigma_1 & 0 & 0 & \ldots & 0 \\
0 & \sigma_2 & 0 & \ldots & 0 \\
0 & 0 & \sigma_3 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & \sigma_n \\
0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 0
\end{bmatrix} V^T
\]

Here \( U \) and \( V \) are orthonormal matrices \((U^T U = I, \ V^T V = I)\) of dimensions \( pr \times pr \) and \( n^* \times n^* \), respectively. \( S \) is a \( pr \times n^* \) matrix with the singular values of \( G \) along the diagonal and zeros elsewhere. If \( G \) has rank \( n \), only the first \( n \) singular values \( \sigma_k \) will be non-zero. This means that we can rewrite

\[
G = USV^T = U_1 S_1 V_1^T
\]

where \( U_1 \) is a \( pr \times n \) matrix containing the first \( n \) columns of \( U \), while \( S_1 \) is the \( n \times n \) upper left part of \( S \), and \( V_1 \) consists of the first \( n \) columns of \( V \).
Subspace Methods for Estimating State Space Models

\[ G = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 2 & 10 \\ 6 & 7 & 19 \\ 1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} -0.19 & -0.40 & 0.90 & -0.03 \\ -0.44 & 0.85 & 0.28 & -0.07 \\ -0.87 & -0.35 & -0.34 & -0.11 \\ -0.14 & 0.01 & 0.00 & 0.99 \end{bmatrix} \begin{bmatrix} 24.39 & 0 & 0 \\ 0 & 1.81 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.30 & 0.49 & 0.82 \\ -0.31 & -0.86 & 0.41 \\ -0.91 & 0.13 & -0.41 \end{bmatrix} \]

\[ G = U S V^T \]

\[ \begin{bmatrix} -0.19 & -0.40 \\ -0.44 & 0.85 \\ -0.87 & -0.35 \\ -0.14 & 0.01 \end{bmatrix} \begin{bmatrix} 24.39 & 0 \\ 0 & 1.81 \end{bmatrix} \begin{bmatrix} -0.30 & 0.49 \\ -0.31 & -0.86 \\ -0.91 & 0.13 \end{bmatrix}^T \]

\[ G = U S V^T = U_1 S_1 V_1^T \]
Subspace Methods for Estimating State Space Models

\[ G = U S V^T = U_1 S_1 V_1^T \]
\[ \tilde{O_r} \tilde{T} = U_1 S_1 V_1^T \]

Now multiplying this by \( V_1 \) from right.

\[ O_r \tilde{T} V_1 = O_r T = U_1 S_1 \]

Now multiplying this by \( S_1^{-1} \) from right.

\[ \hat{O_r} = U_1 \]

Or for some invertible matrix \( R \):

\[ O_r = U_1 R \]

\[ \hat{C} = O_r (1 : p, 1 : n) \]

\[ O_r (p + 1 : pr, 1 : n) = O_r (1 : p(r - 1), 1 : n) \hat{A} \]
Subspace Methods for Estimating State Space Models

Using a Noisy Estimate of the Extended Observability Matrix

Let us now assume that the given $pr \times n^*$ matrix $G$ is a noisy estimate of the true observability matrix

$$ G = O_r \tilde{T} + E_N $$

Where $E_N$ is small and tends to zero as $N \to \infty$.

The rank of $O_r$ is not known. While the noise matrix $E_N$ is likely to be full rank.

It is reasonable to proceed as above and perform an SVD on $G$:

$$ G = USV^T $$

Due to the noise, $S$ will typically have all singular non-zero values

$$ \sigma_k; k = 1, \ldots, \min(n^*, pr) $$
**Subspace Methods for Estimating State Space Models**

The first $n$ will be supported by $O_r$, while the remaining ones will stem from $E_N$.

If the noise is small, one should expect that the latter are significantly smaller than the former.

Therefore determine $n$ as the number of singular values that are significantly larger than 0.

Then use $O_r$ to determine $\hat{A}$, $\hat{C}$ as before. However in the noisy case, $\hat{O}_r$ will not be exactly subject to the shift structure.

So this system of equations should be solved in a least-squares sense.
Subspace Methods for Estimating State Space Models

Using Weighting Matrices in the SVD

For more flexibility we could pre- and post- multiply $G$ as $\hat{G} = W_1GW_2$ before performing the SVD

$$\hat{G} = W_1GW_2 = USV^T \approx U_1S_1V_1^T$$

And then use the below equation to determine $\hat{A}$ and $\hat{C}$

$$\hat{O}_r = W_1^{-1}U_1R$$

Here $R$ is an arbitrary matrix, that will the coordinate basis for the state representation.

The post-multiplication by $W_2$ just corresponds to a change of basis in the state-space and the pre-multiplication by $W_1$ is eliminated.

In the noiseless case $E=0$, these weightings are without consequence. However, when noise is present, they have an important influence on the space spanned by $U_1$ and hence on the quality of the estimates $\hat{A}$ and $\hat{C}$.

Remark. The post-multiplying $W_2$ by an orthogonal matrix does not effect the $U_1$-matrix in the decomposition.

Subspace Methods for Estimating State Space Models

 ► Estimating the Extended Observability Matrix.

Remember

\[ x(t + 1) = Ax(t) + Bu(t) + w(t) \]
\[ y(t) = Cx(t) + Du(t) + v(t) \]

Now,

\[ y(t + k) = Cx(t + k) + Du(t + k) + v(t + k) \]
\[ = CAx(t + k - 1) + CBu(t + k - 1) + Cw(t + k - 1) \]
\[ + Du(t + k) + v(t + k) \]
\[ = \ldots \]
\[ = CA^k x(t) + CA^{k-1} Bu(t) + CA^{k-2} Bu(t + 1) + \ldots \]
\[ + CBu(t + k - 1) + Du(t + k) \]
\[ + CA^{k-1} w(t) + CA^{k-2} w(t + 1) + \ldots \]
\[ + Cw(t + k - 1) + v(t + k) \]
Subspace Methods for Estimating State Space Models

Estimating the Extended Observability Matrix.

\[
y(t + k) = CA^k x(t) + CA^{k-1} Bu(t) + CA^{k-2} Bu(t + 1) + \ldots \\
+ CBu(t + k - 1) + Du(t + k) \\
+ CA^{k-1} w(t) + CA^{k-2} w(t + 1) + \ldots \\
+ Cw(t + k - 1) + \nu(t + k)
\]

Now, form the vectors

\[
Y_r(t) = \begin{bmatrix}
y(t) \\
y(t + 1) \\
\vdots \\
y(t + r - 1)
\end{bmatrix}, \quad U_r(t) = \begin{bmatrix}
u(t) \\
u(t + 1) \\
\vdots \\
u(t + r - 1)
\end{bmatrix}
\]
Subspace Methods for Estimating State Space Models

Estimating the Extended Observability Matrix.

\[ y(t + k) = CA^k x(t) + CA^{k-1} Bu(t) + CA^{k-2} Bu(t + 1) + \ldots \]
\[ + CB u(t + k - 1) + Du(t + k) \]
\[ + CA^{k-1} w(t) + CA^{k-2} w(t + 1) + \ldots \]
\[ + C w(t + k - 1) + v(t + k) \]

\[ Y_r(t) = O_r x(t) + S_r U_r(t) + V(t) \]

\[ O_r = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix} \]

\[ S_r = \begin{bmatrix} D & 0 & \ldots & 0 & 0 \\ CB & D & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ CA^{r-2} B & CA^{r-3} B & \ldots & CB & D \end{bmatrix} \]

And the Kth block component of \( V(t) \)

\[ V_k(t) = CA^{k-2} w(t) + CA^{k-3} w(t + 1) + \ldots + C w(t + k - 2) + v(t + k - 1) \]
Subspace Methods for Estimating State Space Models

勾勒 the Extended Observability Matrix.

\[ Y_r(t) = \begin{bmatrix} y(t) \\ y(t + 1) \\ \vdots \\ y(t + r - 1) \end{bmatrix} \]

\[ U_r(t) = \begin{bmatrix} u(t) \\ u(t + 1) \\ \vdots \\ u(t + r - 1) \end{bmatrix} \]

\[ Y_r(t) = O_r x(t) + S_r U_r(t) + V(t) \]

\[ Y = \begin{bmatrix} Y_r(1) & Y_r(2) & \ldots & Y_r(N) \end{bmatrix} \]

\[ X = \begin{bmatrix} x(1) & x(2) & \ldots & x(N) \end{bmatrix} \]

\[ U = \begin{bmatrix} U_r(1) & U_r(2) & \ldots & U_r(N) \end{bmatrix} \]

\[ V = \begin{bmatrix} V(1) & V(2) & \ldots & V(N) \end{bmatrix} \]

We must eliminate the U term and make the noise influence disappear asymptotically.
Subspace Methods for Estimating State Space Models

Estimating the Extended Observability Matrix.

\[ Y = O_r X + S_r U + V \]

We must eliminate the U term and make the noise influence disappear asymptotically.

**Removing the U-term.** Form the \( N \times N \) matrix

\[
\Pi_{UT}^\perp = I - U^T(UU^T)^{-1}U
\]

Multiplying from the right by \( \Pi_{UT}^\perp \) will leads to:

\[
Y \Pi_{UT}^\perp = O_r X \Pi_{UT}^\perp + V \Pi_{UT}^\perp
\]

\( \checkmark \quad \checkmark \quad \text{?} \)

Since this term is made up of noise contributions, the idea is to correlate is away with a suitable matrix.
Subspace Methods for Estimating State Space Models

- Estimating the Extended Observability Matrix.

\[ Y \Pi_{UT}^{\perp} = O_r X \Pi_{UT}^{\perp} + V \Pi_{UT}^{\perp} \]

Removing the Noise Term. Since the last term is made up of noise contributions. The idea is to correlate it away with a suitable matrix. Define a matrix $(s \geq n)$.

\[ \Phi = [\varphi_s(1) \varphi_s(2) \ldots \varphi_s(N)] \]

\[ \frac{1}{N} Y \Pi_{UT}^{\perp} \Phi^T = O_r \frac{1}{N} X \Pi_{UT}^{\perp} \Phi^T + \frac{1}{N} V \Pi_{UT}^{\perp} \Phi^T \]

Here $\Phi$ acts as an instrument and we must define it such that

\[
\lim_{N \to \infty} V_N = \lim_{N \to \infty} \frac{1}{N} V \Pi_{UT}^{\perp} \Phi^T = 0 \quad \lim_{N \to \infty} \tilde{T}_N = \lim_{N \to \infty} \frac{1}{N} X \Pi_{UT}^{\perp} \Phi^T = \tilde{T} \text{ has full rank } n
\]
Subspace Methods for Estimating State Space Models

Estimating the Extended Observability Matrix.

\[
\frac{1}{N} Y \Pi_{UT}^{+} \Phi^T = O_r \frac{1}{N} X \Pi_{UT}^{+} \Phi^T + \frac{1}{N} V \Pi_{UT}^{+} \Phi^T
\]

Here \( \Phi \) acts as an instrument and we must define it such that

\[
\lim_{N \to \infty} V_N = \lim_{N \to \infty} \frac{1}{N} Y \Pi_{UT}^{+} \Phi^T = 0 \quad \lim_{N \to \infty} \tilde{T}_N = \lim_{N \to \infty} \frac{1}{N} X \Pi_{UT}^{+} \Phi^T = \tilde{T} \quad \text{has full rank } n
\]

then

\[
G \triangleq O_r \tilde{T}_N + V_N = O_r \tilde{T} - O_r \tilde{T} + O_r \tilde{T}_N + V_N
\]

so

\[
G = \frac{1}{N} Y \Pi_{UT}^{+} \Phi^T = O_r \tilde{T} + E_N
\]

\[
E_N = O_r (\tilde{T}_N - \tilde{T}) + V_N \to 0 \text{ as } N \to \infty
\]

The \( pr \times s \) matrix \( G \) can thus be seen as a noisy estimate of the extended observability matrix.

But we need to define \( \Phi \).
Finding Good Instruments. The only remaining question is how to achieve to the following equations

\[
\lim_{N \to \infty} V_N = \lim_{N \to \infty} \frac{1}{N} V \Pi_{U^T} \Phi^T = 0 \quad \lim_{N \to \infty} \tilde{T}_N = \lim_{N \to \infty} \frac{1}{N} X \Pi_{U^T} \Phi^T = \tilde{T} \quad \text{has full rank } n
\]

Remember instrument variable:

\[
\bar{E} \zeta(t) v_0(t) = 0 \quad \bar{E} \zeta(t) \varphi^T(t) \text{ be nonsingular}
\]

\[
\frac{1}{N} V \Pi_{U^T} \Phi^T = \frac{1}{N} \sum_{t=1}^{N} V(t) \varphi_s^T(t) - \frac{1}{N} \sum_{t=1}^{N} V(t) U_r^T(t) \times \left[ \frac{1}{N} \sum_{t=1}^{N} U_r(t) U_r^T(t) \right]^{-1} \frac{1}{N} \sum_{t=1}^{N} U_r(t) \varphi_s^T(t)
\]

Remember:

\[
\Pi_{U^T} = I - U^T (UU^T)^{-1} U
\]

The law of large numbers states that the sample sums converges to their respective expected values, so

\[
\lim_{N \to \infty} \frac{1}{N} V \Pi_{U^T} \Phi^T = \bar{E} V(t) \varphi_s^T(t) - \bar{E} V(t) U_r^T(t) R_u^{-1} \bar{E} U_r(t) \varphi_s^T(t)
\]
Subspace Methods for Estimating State Space Models

Finding Good Instruments. The only remaining question is how to achieve to the following equations

\[
\lim_{N \to \infty} V_N = \lim_{N \to \infty} \frac{1}{N} V \Pi_{U^T} \Phi^T = 0 \quad \lim_{N \to \infty} \tilde{T}_N = \lim_{N \to \infty} \frac{1}{N} X \Pi_{U^T} \Phi^T = \tilde{T} \quad \text{has full rank } n
\]

\[
\lim_{N \to \infty} \frac{1}{N} V \Pi_{U^T} \Phi^T = \overline{E} V(t) \varphi_s^T(t) - \overline{E} V(t) U_r^T(t) R_u^{-1} \overline{E} U_r(t) \varphi_s^T(t)
\]

Assume the input \( u \) is generated in open loop, so that it is independent of the noise \( V \).

\[
V_k(t) = CA^{k-2}w(t) + CA^{k-3}w(t + 1) + \ldots + Cw(t + k - 2) + \nu(t + k - 1)
\]

Now since \( V(t) \) is made of white noise term from time \( t \) and onwards so:

\[
\varphi_s(t) = \begin{bmatrix} y(t - 1) \\
\vdots \\
y(t - s_1) \\
u(t - 1) \\
\vdots \\
u(t - s_2) \end{bmatrix}
\]
Finding Good Instruments. The only remaining question is how to achieve to the following equations

\[
\lim_{N \to \infty} V_N = \lim_{N \to \infty} \frac{1}{N} \mathbf{V} \mathbf{P}^{-1}_{U_T} \mathbf{P}^T = 0 \quad \lim_{N \to \infty} \tilde{T}_N = \lim_{N \to \infty} \frac{1}{N} \mathbf{X} \mathbf{P}^{-1}_{U_T} \mathbf{P}^T = \tilde{T} \text{ has full rank } n
\]

Similarly we have:

\[
\tilde{T} = \overline{E} x(t) \varphi^T_S(t) - \overline{E} x(t) U^T_r(t) R_u^{-1} \overline{E} U_r(t) \varphi^T_S(t)
\]

A formal proof that \( \tilde{T} \) has full rank is not immediate and will involve properties of the input.

Problem 10G.6 show the suitable input.
Subspace Methods for Estimating State Space Models

Finding the States and Estimating the Noise statistics

Some part of chapter 7

Let a system given by the impulse response representation

$$y(t) = \sum_{j=0}^{\infty} h_u(j)u(t-j) + h_e(j)e(t-j) \quad (I)$$

Formal k-step ahead predictors be defined:

$$\hat{y}(t \mid t-k) = \sum_{j=k}^{\infty} h_u(j)u(t-j) + h_e(j)e(t-j)$$

Define

$$\hat{Y}_r(t) = \begin{bmatrix} \hat{y}(t \mid t-1) \\ \vdots \\ \hat{y}(t+r-1 \mid t-1) \end{bmatrix}$$

$$\hat{Y} = \begin{bmatrix} \hat{Y}_r(1) & \cdots & \hat{Y}_r(N) \end{bmatrix}$$

Then the following is true as $N \to \infty$ (see chapter 4 appendix A)

1- The system (I) has an $n$th order minimal state space description if and only if the rank $\hat{Y}$ is equal to $n$ for all $r \geq n$

2- The state vector of any minimal realizations form can be chosen as linear

$$x(t) = L\hat{Y}_r(t)$$
Subspace Methods for Estimating State Space Models

Finding the States and Estimating the Noise statistics

Let a system given by the impulse response representation

\[ \hat{y}(t | t-k) = \sum_{j=k}^{\infty} h_u(j)u(t-j) + h_e(j)e(t-j) \]
\[ \hat{y}(t+k-1 | t-1) = \sum_{j=1}^{\infty} h_u(j)u(t-j) + h_e(j)e(t-j) \]

For practical reason we have

\[ \hat{y}(t+k-1 | t-1) = \alpha_1y(t-1) + \ldots + \alpha_{s_1}y(t-s_1) + \beta_1u(t-1) + \ldots + \beta_{s_2}u(t-s_2) \]

This predictor can be determined effectively by

\[ y(t+k-1) = \theta_k^T \varphi_s(t) + \gamma_k^T U_l(t) + \varepsilon(t+k-1) \]

or, dealing with all \( r \) predictors simultaneously

\[ Y_r(t) = \Theta \varphi_s(t) + \Gamma U_r(t) + E(t) \]

\[ Y = [Y_r(1) \ Y_r(2) \ldots \ Y_r(N)] \]

\[ U = [U_r(1) \ U_r(2) \ldots \ U_r(N)] \]
Subspace Methods for Estimating State Space Models

Finding the States and Estimating the Noise statistics

\[ Y = \Theta \Phi + \Gamma U + E \]

By LS we have

\[
[\hat{\Theta} \, \hat{\Gamma}] = [Y \Phi^T \, Y U^T] \left[ \begin{array}{cc} \Phi \Phi^T & \Phi U^T \\ U \Phi^T & U U^T \end{array} \right]^{-1}
\]

By inverse lemma

\[
\hat{\Theta} = Y \Pi_{UT}^{-1} \Phi^T (\Phi \Pi_{UT}^{-1} \Phi^T)^{-1}
\]

\[
\hat{\Gamma} = W_2 \left( \Phi \Pi_{UT}^{-1} \Phi^T \right)^{-1} \Phi
\]

Remember

\[
\frac{1}{N} \Pi_{UT} \Phi^T = O_r \left( \frac{1}{N} X \Pi_{UT} \Phi^T + \frac{1}{N} V \Pi_{UT} \Phi^T \right)
\]

\[
\hat{G} = W_1 G W_2
\]

So we have

\[
\hat{Y} \approx U_1 S_1 V_1^T
\]

\[
O_r = U_1 R
\]

\[
\hat{Y} \approx U_1 R_1 R_1^{-1} S_1 V_1^T
\]
Finding the States and Estimating the Noise statistics

\[ Y = \Theta \Phi + \Gamma U + E \]

\[ \hat{Y} \approx U_1 R_1 R_1^{-1} S_1 V_1^T \]

\[ \hat{Y} \approx U_1 R_1 R_1^{-1} S_1 V_1^T = O_r \hat{X} \quad O_r = U_1 R \]

Remember

\[ Y = O_r X + S_r U + V \]

So let:

\[ \hat{Y} = O_r \hat{X} \quad \hat{X} = O_r^{-1} \hat{Y} \quad \hat{X} = R^{-1} U_1^T \hat{Y} = [\hat{x}(1) \ldots \hat{x}(N)] \]

With the states given, we can estimate the process and measurement noises as

\[ w(t) = \hat{x}(t + 1) - \hat{A} \hat{x}(t) - \hat{B} u(t) \]

\[ v(t) = y(t) - \hat{C} \hat{x}(t) - \hat{D} u(t) \]
Subspace Methods for Estimating State Space Models

Putting It All Together

The family of subspace algorithm

1. From the input-output data form

\[ G = \frac{1}{N} Y \Pi_{U^T} \Phi^T \]

Remember:

\[ Y_r(t) = \begin{bmatrix} y(t) \\ y(t + 1) \\ \vdots \\ y(t + r - 1) \end{bmatrix} \quad U_r(t) = \begin{bmatrix} u(t) \\ u(t + 1) \\ \vdots \\ u(t + r - 1) \end{bmatrix} \]

\[ \Pi_{U^T} = I - U^T(UU^T)^{-1}U \]

\[ Y = [Y_r(1) \quad Y_r(2) \quad \ldots \quad Y_r(N)] \]

\[ U = [U_r(1) \quad U_r(2) \quad \ldots \quad U_r(N)] \]

Many algorithms choose \( \phi_s(t) \) to consist of past inputs and outputs with \( s_1 = s_2 = s \). So scalar \( s \) is a design variable.

The scalar \( r \), is the maximal prediction horizon and in many algorithms use \( r = s \).
Subspace Methods for Estimating State Space Models

Putting It All Together

The family of subspace algortithm

2. Select weighting matrices $W_1$ and $W_2$ and perform SVD

$\hat{G} = W_1 G W_2 = U S V^T \approx U_1 S_1 V_1^T$

The weighting matrices $W_1$ and $W_2$. This is the perhaps most important choice. Existing algorithms employ the following choices:

- **MOESP, Verhaegen (1994):** $W_1 = I$, $W_2 = \left(\frac{1}{N} \Phi \Pi_{UT} \Phi^T\right)^{-1} \Phi \Pi_{UT}$

- **N4SID, Van Overschee and DeMoor (1994):** $W_1 = I$, $W_2 = \left(\frac{1}{N} \Phi \Pi_{UT} \Phi^T\right)^{-1} \Phi$ (see also (10.120)).

- **IVM, Viberg (1995):** $W_1 = \left(\frac{1}{N} Y \Pi_{UT} Y\right)^{-1/2}$, $W_2 = \left(\frac{1}{N} \Phi \Phi^T\right)^{-1/2}$

- **CVA, Larimore (1990):** $W_1 = \left(\frac{1}{N} Y \Pi_{UT} Y\right)^{-1/2}$, $W_2 = \left(\frac{1}{N} \Phi \Pi_{UT} \Phi^T\right)^{-1/2}$
Subspace Methods for Estimating State Space Models

Putting It All Together

The family of subspace algotithm

3. Select a full rank matrix $R$ and define the $r p \times n$ matrix $\hat{O}_r = W_1^{-1} U_1 R$ solve

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

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

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

For $\hat{C}$ and $\hat{A}$. The latter equation should be solved in a least square sense.

4. Estimate $\hat{B}$, $\hat{D}$ and $\hat{x}_0$ from the linear regression problem:

$$\arg\min_{B, D, x_0} \frac{1}{N} \sum_{t=1}^{N} \left\| y(t) - \hat{C} (q I - \hat{A})^{-1} B u(t) - D u(t) - \hat{C} (q I - \hat{A})^{-1} x(t) \delta(t) \right\|^2$$
The family of subspace algorithm

5. If a noise model is sought, form $\hat{X}$ as in

$$\hat{X} = L\hat{Y} = \begin{bmatrix} \hat{x}(1) & \ldots & \hat{x}(N) \end{bmatrix}$$

$$L = R^{-1}U_j^T$$

And estimate the noise contributions as in

$$w(t) = \hat{x}(t + 1) - \hat{A}\hat{x}(t) - \hat{B}u(t)$$

$$v(t) = y(t) - \hat{C}\hat{x}(t) - \hat{D}u(t)$$