State-Space Identification, Sub-space Methods & Instrumental-Variable Methods

Jonas Sjöberg

May 11, 2020
State-Space Models

- Linear state space models
- MIMO state space models
- Grey-box modelling as continuous time state space models
- Computing parameter estimates
  - PEM
  - Instrumental-Variable Methods (IV)
  - Subspace Algorithms
Linear State-Space Models

System:

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

Input signals: \( u(t) \)  
Output signals: \( y(t) \)  
State vector: \( x(t) \)  
Process noise: \( w(t) \)  
Measurement noise: \( v(t) \)

Model (innovation form):

\[
\begin{align*}
\hat{x}(t + 1, \theta) &= A(\theta)\hat{x}(t, \theta) + B(\theta)u(t) + K(\theta)e(t) \\
y(t) &= C(\theta)\hat{x}(t, \theta) + D(\theta)u(t) + e(t)
\end{align*}
\]

Compare Kalman filter, \( e(t) \to \epsilon(t) \), \( \hat{y}(t | t - 1) = C \hat{x}(t) + D u(t) \)
State space form is just an alternative way to express the plant and noise model:

\[
\begin{align*}
    y(t) &= G(q, \theta)u(t) + H(q, \theta)e(t) \\
    G(q, \theta) &= C(\theta)[qI - A(\theta)]^{-1}B(\theta) \\
    H(q, \theta) &= C(\theta)[qI - A(\theta)]^{-1}K(\theta) + I
\end{align*}
\]
Different parametrizations, eg observable canonical form:

\[ A(\theta) = \begin{pmatrix} -a_1 & 1 & 0 \\ -a_2 & 0 & 1 \\ -a_3 & 0 & 0 \end{pmatrix} \]

\[ B(\theta) = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \quad K(\theta) = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} \]

\[ C(\theta) = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \quad D(\theta) = 0 \]

\[ G(q, \theta) = C(\theta)[qI - A(\theta)]^{-1}B(\theta) = \frac{b_1q^{-1} + b_2q^{-2} + b_3q^{-3}}{1 + a_1q^{-1} + a_2q^{-2} + a_3q^{-3}} \]

\[ H(q, \theta) = C(\theta)[qI - A(\theta)]^{-1}K(\theta) + 1 = \frac{k_1q^{-1} + k_2q^{-2} + k_3q^{-3}}{1 + a_1q^{-1} + a_2q^{-2} + a_3q^{-3}} + 1 \]
Example: 2 inputs, 3 outputs, 8 states

\[
A(\theta) = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\times & \times & \times & \times & \times & \times & \times & \times \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\times & \times & \times & \times & \times & \times & \times & \times \\
\end{pmatrix}
\]

\[
B(\theta) = \begin{pmatrix}
\times & \times \\
\vdots & \vdots \\
\times & \times \\
\end{pmatrix} \quad K(\theta) = \begin{pmatrix}
\times & \times & \times \\
\vdots & \vdots & \vdots \\
\times & \times & \times \\
\end{pmatrix}
\]

\[
C(\theta) = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
\end{pmatrix}
\]
Problem:
- Rows for the parameters in the A-matrix must be chosen.
- Bad choice gives ill-conditioned computations when the estimate is computed.
- This problem is avoided if subspace methods are used, they have full matrices.

\[
A(\theta) = \\
\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\times & \times & \times & \times & \times & \times & \times & \times \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\times & \times & \times & \times & \times & \times & \times & \times \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\times & \times & \times & \times & \times & \times & \times & \times \\
\end{pmatrix}
\]
Grey-box modelling (physical modelling, first-principle modelling)

If the model structure is based on physical modelling but with unknown parameters to be identified, then it is often natural to do the modelling with a time-continuous state space model.

\[
\begin{align*}
\dot{x}(t) &= f(\theta, x(t), u(t), \varepsilon(t)) \\
\hat{y}(t) &= g(\theta, x(t), u(t))
\end{align*}
\]

A discrete-time model can be obtained by using Euler forward

\[
\begin{align*}
x(t + T) &= x(t) + T \cdot f(\theta, x(t), u(t), \varepsilon(t)) \\
\hat{y}(t) &= g(\theta, x(t), u(t))
\end{align*}
\]

- Tools for this are available, eg, in SITB.
- Linear or nonlinear model
Computing parameter estimates

- PEM
- Instrumental Variable Method
- Subspace Methods
Any model can be re-written on state space form, hence, consider state space model (linear or nonlinear)

\[
x(t + 1) = f(\theta, x(t), u(t), \varepsilon(t)) \\
\hat{y}(t) = g(\theta, x(t), u(t))
\]

\( \hat{\theta} \) defined as \( \hat{\theta}_N = \arg \min_{\theta} V_N(\theta) \) where

\[
V_N(\theta) = \frac{1}{N} \sum_{t} \varepsilon^2(\theta, t)
\]

and \( \varepsilon(\theta, t) = y(t) - \hat{y}(\theta, t) \).
Search for the minimum of $V_N(\theta)$:

Given an initial estimate $\theta^0$, iterate

$$
\theta^{(i+1)} = \theta^{(i)} - \mu_i R_i \frac{dV_N(\theta)}{d\theta}
$$

until the minimum is reached. $\mu_i$ is a step length to assure downhill steps, and $R_i$ is a matrix to modify the search direction.

Note: a lot of research on system identification concerns methods to obtain a good initial value $\theta^0$ so obtain better chances for convergence to global minimum.
\[
\frac{dV_N(\theta)}{d\theta} = -\frac{2}{N} \sum_{t=1}^{N} \varepsilon(\theta, t) \frac{d\hat{y}(t)}{d\theta}
\]

The derivative of model output

\[
\frac{d\hat{y}(t)}{d\theta} = \frac{dg(\theta, x(t), u(t))}{d\theta} = \frac{\partial g(\theta, x(t), u(t))}{\partial \theta} + \frac{\partial g(\theta, x(t), u(t))}{\partial x(t)} \cdot \frac{dx(t)}{d\theta}
\]

where

\[
\frac{dx(t)}{d\theta}
\]

given by ....
... where

\[
\frac{dx(t)}{d\theta}, \quad x(t + 1) = f(\theta, x(t), u(t), \varepsilon(t))
\]

given by

\[
\frac{dx(t + 1)}{d\theta} = \frac{\partial f(\theta, x(t), u(t), \varepsilon(t))}{\partial x(t)} \cdot \frac{dx(t)}{d\theta} + \frac{\partial f(\theta, x(t), u(t), \varepsilon(t))}{\partial \theta}
\]

this is (nonlinear) filtering with input signal

\[
\frac{\partial f(\theta, x(t), u(t), \varepsilon(t))}{\partial \theta}.
\]

Depends, however, on \(x(t)\) which, must first be obtained by using the model itself.
Linear first order OE model

\begin{equation}
  x(t) = \theta_1 x(t - 1) + \theta_2 u(t), \quad \hat{y}(t) = x(t)
\end{equation}

The derivative with respect \( \theta_1 \)

\begin{equation}
  \frac{dx(t)}{d\theta_1} = \theta_1 \frac{dx(t - 1)}{d\theta_1} + x(t - 1).
\end{equation}

This is a filter with input \( x(t) \) and output

\[ \frac{dx(t)}{d\theta_1}. \]

Setup PEM:

- Initial estimate
- Iterative minimization
Instrumental-Variable Methods

- Estimate plant model without iterative minimization (as for PEM). Avoid bias (as for ARX model).
- Something like LS but without the bias problem.
- No disturbance model (how can that be fixed?)

Consider data generated by

\[ y(t) = \varphi^T(t)\theta_0 + v_0(t) \]

The LS solution can be expressed as (the normal equation)

\[ \hat{\theta}_{LS}^N = \text{sol} \left\{ \frac{1}{N} \sum_{t=1}^{N} \varphi(t)[y(t) - \varphi^T(t)\theta] = 0 \right\} \]

In general \( \hat{\theta}_{LS}^N \to \theta_0 \) when \( v(t) \) and \( \varphi(t) \) correlated, eg, a ARX model.
Idea: replace $\varphi(t)$ with $\xi(t)$ uncorrelated with $v(t)$.

$$\hat{\theta}^{IV}_N = \text{sol} \left\{ \frac{1}{N} \sum_{t=1}^{N} \xi(t)[y(t) - \varphi^T(t)\theta] = 0 \right\}$$

or

$$\hat{\theta}^{IV}_N = \left[ \frac{1}{N} \sum_{t=1}^{N} \xi(t)\varphi^T(t) \right]^{-1} \frac{1}{N} \sum_{t=1}^{N} \xi(t)y(t)$$

The elements of $\xi(t)$ are the *instruments*, or *instrumental variable*.

Requirements:

- $\bar{E}\xi(t)\varphi^T(t)$ nonsingular
- $\bar{E}\xi(t)v_0(t) = 0$

$\xi(t)$ is obtained by filtering $u(t)$. Optimal filter is the true (unknown) system...
\( \xi(t) \) is obtained by filtering \( u(t) \). Optimal filter is the true (unknown) system.

\[
y(t) = \hat{y}(t) + e(t) \quad e(t) \text{ not necessary white}
\]

\[
\phi^T(t) = [-y(t-1) \ - y(t-2) \ldots] \quad \xi^T(t) = [-\hat{y}(t-1) \ - \hat{y}(t-2) \ldots]
\]

\[
\hat{\theta}_{IV}^N = \left[ \frac{1}{N} \sum_{t=1}^{N} \xi(t)\phi^T(t) \right]^{-1} \frac{1}{N} \sum_{t=1}^{N} \xi(t)y(t) \approx \left[ \frac{1}{N} \sum_{t=1}^{N} \xi(t)\xi^T(t) \right]^{-1} \frac{1}{N} \sum_{t=1}^{N} \xi(t)\hat{y}(t)
\]

This is what you did in project 2, but explained slightly different. Inserting

\[
y(t) = \phi(t)\theta_0 + e(t)
\]

verifies that the asymptotic estimate is unbiased.
A state-space model

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

is not a linear regression model. Drawbacks using PE:

- can be hard to choose parametrization, especially for MIMO models.
- iterative algorithm.
- often ill-conditioned.
\[
x(t + 1) = Ax(t) + Bu(t) + w(t) \\
y(t) = Cx(t) + Du(t) + v(t)
\]

Subspace algorithms:
- The estimate is obtained as a series of projections. First is \( x(t) \) obtained, then parameters are obtained by LS.
- Can be interpreted as instrumental variable methods
- No iterations – much faster than PE.
- No criterion is minimized.
- There are situations where subspace algorithms fails.