System Identification

Control Engineering EN, 3rd year B.Sc.
Technical University of Cluj-Napoca
Romania

Lecturer: Lucian Buşoniu

Part VI

Instrumental variable methods

Table of contents

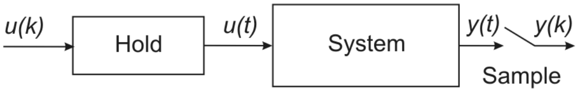
Classification

1. Mental or verbal models
2. Graphs and tables (nonparametric)
3. Mathematical models, with two subtypes:
   - First-principles, analytical models
   - Models from system identification

Like prediction error methods, instrumental variable methods produce *parametric*, polynomial models.

Table of contents

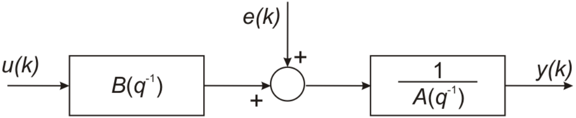
Setting: discrete time SISO



We stay in the single-output, single-input case in this part.

Recall ARX model

$$A(q^{-1})y(k) = B(q^{-1})u(k) + e(k)$$
$$(1+a_1q^{-1} + \cdots + a_{na}q^{-na})y(k) =$$
$$(b_1q^{-1} + \cdots + b_{nb}q^{-nb})u(k) + e(k)$$



In explicit form:

$$y(k) + a_1y(k-1) + a_2y(k-2) + \ldots + a_{na}y(k-na)$$
$$= b_1u(k-1) + b_2u(k-2) + \ldots + b_{nb}u(k-nb) + e(k)$$

where the model parameters are: $a_1, a_2, \ldots, a_{n_a}$ and $b_1, b_2, \ldots, b_{nb}$.

Linear regression representation

$$
\begin{aligned}
y(k) = &- a_1 y(k-1) - a_2 y(k-2) - \ldots - a_{na} y(k-na) \\
&\quad b_1 u(k-1) + b_2 u(k-2) + \ldots + b_{nb} u(k-nb) + e(k) \\
= &\begin{bmatrix} -y(k-1) & \cdots & -y(k-na) & u(k-1) & \cdots & u(k-nb) \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} a_1 & \cdots & a_{na} & b_1 & \cdots & b_{nb} \end{bmatrix}^\top + e(k) \\
=: &\varphi^\top(k)\theta + e(k)
\end{aligned}
$$

Regressor vector: $\varphi \in \mathbb{R}^{na+nb}$, previous output and input values.

Parameter vector: $\theta \in \mathbb{R}^{na+nb}$, polynomial coefficients.

Identification problem and solution

Given dataset $u(k), y(k)$, $k = 1, \dots, N$, find model parameters θ to achieve small equation errors $\varepsilon(k)$ in:

$$y(k) = \varphi^\top(k)\theta + \varepsilon(k)$$

Formal objective: minimize the mean squared error:

$$V(\theta) = \frac{1}{N} \sum_{k=1}^{N} \varepsilon(k)^2$$

Solution: can be written in several ways, we use:

$$\widehat{\theta} = \left[\frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^\top(k) \right]^{-1} \left[\frac{1}{N} \sum_{k=1}^{N} \varphi(k)y(k) \right]$$

Theoretical guarantee

Recall that for the guarantees, a true parameter vector θ_0 is assumed to exist:

$$y(k) = \varphi^\top(k)\theta_0 + v(k)$$

Analyze the parameter errors (a vector of n elements):

$$\begin{aligned}
\widehat{\theta} - \theta_0 &= \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^\top(k)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)y(k)\right] \\
&\quad - \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^\top(k)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^\top(k)\right]\theta_0 \\
&= \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^\top(k)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)[y(k) - \varphi^\top(k)\theta_0]\right] \\
&= \left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^\top(k)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}\varphi(k)v(k)\right]
\end{aligned}$$

Further assumptions

We wish the algorithm to be consistent: the parameter errors should become 0 in the limit of infinite data (and they should be well-defined).

As $N \to \infty$:

$$\frac{1}{N}\sum_{k=1}^{N}\varphi(k)\varphi^\top(k) \to \mathrm{E}\left\{\varphi(k)\varphi^\top(k)\right\}$$

$$\frac{1}{N}\sum_{k=1}^{N}\varphi(k)v(k) \to \mathrm{E}\left\{\varphi(k)v(k)\right\}$$

For the error to be (1) well-defined and (2) equal to zero, we need:

1. $\mathrm{E}\left\{\varphi(k)\varphi^\top(k)\right\}$ invertible.
2. $\mathrm{E}\left\{\varphi(k)v(k)\right\}$ zero.

Table of contents

Motivation: ARX requires white noise

- We have $\mathrm{E}\{\varphi(k)v(k)\} = 0$ if the elements of $\varphi(k)$ are uncorrelated with $v(k)$ (note that $v(k)$ is assumed zero-mean).
- But $\varphi(k)$ includes $y(k-1), y(k-2), \dots$, which depend on $v(k-1), v(k-2), \dots$!
- So the only option is to have $v(k)$ uncorrelated with $v(k-1), v(k-2), \dots \quad \Rightarrow \quad$ usually, $v(k)$ must be *white noise*.

Instrumental variables are a solution to remove this limitation to white noise.

Intuition

$$\widehat{\theta} - \theta_0 = \left[\frac{1}{N}\sum_{k=1}^{N} \varphi(k)\varphi^\top(k)\right]^{-1} \left[\frac{1}{N}\sum_{k=1}^{N} \varphi(k)v(k)\right]$$

Idea: What if a different vector than $\varphi(k)$ could be included in the parameter errors?

$$\widehat{\theta} - \theta_0 = \left[\frac{1}{N}\sum_{k=1}^{N} Z(k)\varphi^\top(k)\right]^{-1} \left[\frac{1}{N}\sum_{k=1}^{N} Z(k)v(k)\right]$$

where the elements of $Z(k)$ are uncorrelated with $v(k)$. Then $\mathrm{E}\left\{Z(k)v(k)\right\} = 0$ and the error can be zero.

Vector $Z(k)$ has n elements, which are called instruments.

Instrumental variable method

In order to have:

$$\widehat{\theta}-\theta_0=\left[\frac{1}{N}\sum_{k=1}^{N}Z(k)\varphi^\top(k)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}Z(k)v(k)\right] \quad (6.1)$$

the estimated parameter must be:

$$\widehat{\theta}=\left[\frac{1}{N}\sum_{k=1}^{N}Z(k)\varphi^\top(k)\right]^{-1}\left[\frac{1}{N}\sum_{k=1}^{N}Z(k)y(k)\right] \quad (6.2)$$

This $\widehat{\theta}$ is the solution to the system of n equations:

$$\left[\frac{1}{N}\sum_{k=1}^{N}Z(k)[\varphi^\top(k)\theta-y(k)]\right]=0 \quad (6.3)$$

Constructing and solving this system gives the basic instrumental variable (IV) method.

Exercise: Show that (6.3) implies (6.2), and that (6.2) implies (6.1).

Simple instruments

So far the instruments $Z(k)$ were not discussed. They are usually created based on the inputs (including outputs would lead to correlation with v and so eliminate the advantage of IV).

Simple possibility: just include additional delayed inputs to obtain a vector of the appropriate size, $n = na + nb$:

$$Z(k) = [u(k - nb - 1), \ldots u(k - na - nb), u(k - 1), \ldots, u(k - nb)]^\top$$

Compare to original vector:

$$\varphi(k) = [-y(k - 1), \ldots, -y(k - na), u(k - 1), \ldots, u(k - nb)]^\top$$

Generalization

Pass the input through a transfer function:

$$
\begin{aligned}
C(q^{-1})x(k) = D(q^{-1})u(k) \\
(1+c_1 q^{-1} + \dots + c_{nb} q^{-nc})x(k) = \\
(d_1 q^{-1} + \dots + d_{nd} q^{-nd})u(k) \\
x(k) = -c_1 x(k-1) - c_2 x(k-2) - \ldots - c_{nc} x(k-nc) \\
d_1 u(k-1) + d_2 u(k-2) + \ldots + d_{nd} u(k-nd)
\end{aligned}
$$

and take *na* past values from the output x:

$$Z(k) = [-x(k-1), \ldots, -x(k-na), u(k-1), \ldots, u(k-nb)]^\top$$

Remark: $C(q^{-1})$, $D(q^{-1})$ have different meanings than in PEM.

Generalized instruments: obtaining the simple case

In order to obtain:

$$Z(k) = [u(k - nb - 1), \dots u(k - na - nb), u(k - 1), \dots, u(k - nb)]^\top$$

set $C = 1$, $D = -q^{-nb}$.

Exercise: Verify that the desired $Z(k)$ is indeed obtained.

Generalized instruments: Initial model

Generalized instruments:

$$Z(k) = [-x(k-1), \dots, -x(k-na), u(k-1), u(k-2), \dots, u(k-nb)]^\top$$

Compare to original vector:

$$\varphi(k) = [-y(k-1), \dots, -y(k-na), u(k-1), \dots, u(k-nb)]^\top$$

Idea: Take instrument generator equal to an initial model, $C(q^{-1}) = \widehat{A}(q^{-1})$, $D(q^{-1}) = \widehat{B}(q^{-1})$. This model can be obtained e.g. with ARX estimation.

The instruments are an approximation of y:

$$Z(k) = [-\hat{y}(k-1), \dots - \hat{y}(k-na), u(k-1), \dots, u(k-nb)]^\top$$

that is *uncorrelated* with the noise.

Table of contents

Comparison

Both PEM and IV can be seen as extensions of ARX:

$$A(q^{-1})y(k) = B(q^{-1})u(k) + e(k)$$

to disturbances $v(k)$ different from white noise $e(k)$.

- PEM explicitly include the disturbance model in the structure, e.g. in ARMAX $v(k) = C(q^{-1})e(k)$ leading to $A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})e(k)$.
- IV methods do *not* explicitly model the disturbance, but are designed to be resilient to non-white, “colored” disturbance, by using instruments $Z(k)$ uncorrelated with it.

Comparison (continued)

Advantage of IV: Simple model structure, identification consists only of solving a system of linear equations. In contrast, PEM required solving optimization problems with Newton's method, was susceptible to local minima etc.

Disadvantage of IV: In practice, for finite number N of data, model quality depends heavily on the choice of instruments $Z(k)$. Moreover, the resulting model has a larger risk of being unstable (even with a stable real system).

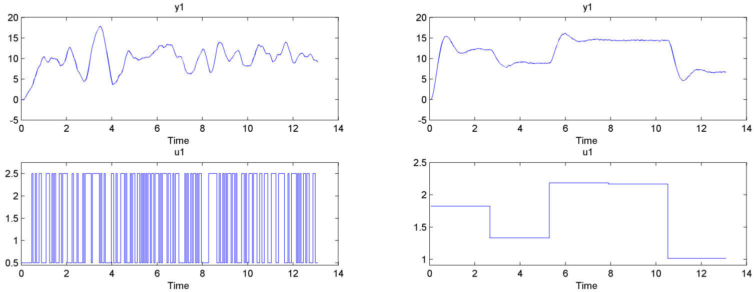
Methods exist to choose instruments $Z(k)$ that are optimal in a certain sense, but they will not be discussed here.

Table of contents

Experimental data

Separate identification and validation data sets:

`plot(id);` and `plot(val);`



From prior knowledge, the system has order 2 and the disturbance is colored (does not obey the ARX model structure).

Remarks: As before, the identification input is a pseudo-random binary signal, and the validation input a sequence of steps.

IV identification with custom instruments

Define the instruments by the generating transfer function, using polynomials $C(q^{-1})$ and $D(q^{-1})$.

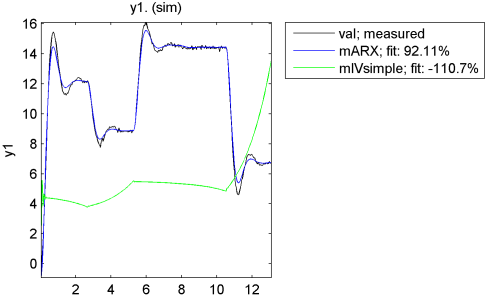
```
model = iv(id, [na, nb, nk], C, D);
```

Arguments:

1. Identification data.
2. Array containing the orders of A and B and the delay nk (like for ARX).
3. Polynomials C and D, as vectors of coefficients in increasing power of q^{-1}.

Result with simple instruments

Take $C(q^{-1}) = 1$, $D(q^{-1}) = -q^{-nb}$, leading to
$Z(k) = [u(k-nb-1), \dots u(k-na-nb), u(k-1), \dots, u(k-nb)]^\top$.
Compare to ARX.



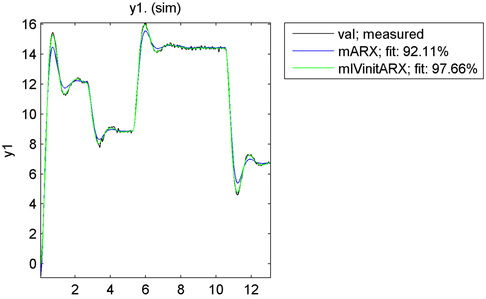
Conclusions:

- Model unstable $\Rightarrow$ in general, must pay attention because IV models are not guaranteed to be stable! (recall the Comparison)
- Results very bad with this simple choice.

Result with ARX-model instruments

Take $C(q^{-1}) = \hat{A}(q^{-1})$, $D(q^{-1}) = \hat{B}(q^{-1})$ from the ARX experiment, leading to

$Z(k) = [-\hat{y}(k-1), \ldots - \hat{y}(k-na), u(k-1), \ldots, u(k-nb)]^\top$.

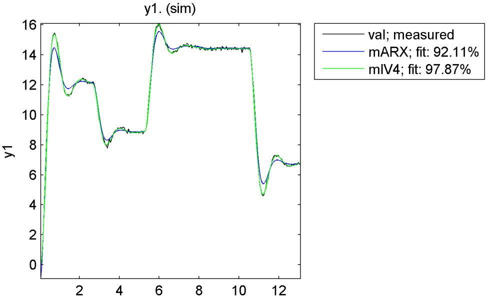


Conclusion: IV obtains better results. This is because the disturbance is colored, and IV can deal effectively with this case (whereas ARX cannot).

Result with automatic instruments

```
model = iv4(id, [na, nb, nk]);
```

Implements an algorithm that generates near-optimal instruments.



Conclusion: Virtually the same performance as ARX instruments.

Table of contents

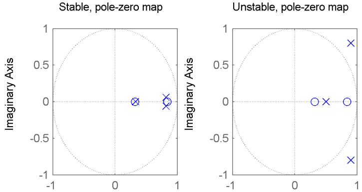
Assumptions

Assumptions (simplified)

1. The disturbance $v(k) = H(q^{-1})e(k)$ where $e(k)$ is zero-mean white noise, $H(q^{-1})$ is a transfer function satisfying certain conditions.
2. The input signal $u(k)$ is "sufficiently informative" and does not depend on the disturbance (the experiment is open-loop).
3. The real system is stable and *uniquely* representable by the model chosen: there exists exactly one θ_0 so that polynomials $A(q^{-1}; \theta_0)$ and $B(q^{-1}; \theta_0)$ are identical to those of the real system.
4. Matrix $\mathrm{E}\left\{Z(k)Z^\top(k)\right\}$ is invertible.

Discussion of assumptions

- Assumption 1 shows the main advantage of IV over PEM: the disturbance can be colored.
- Assumptions 2 and 3 are not very different from those made by PEM. Stability of a discrete-time system requires its poles to be strictly inside the unit circle:



Question: Why is the experiment not allowed to be closed-loop?

- Assumption 4 is required to solve the linear system, and given a sufficiently informative input boils down to an appropriate selection of instruments (e.g. not repeating the same delayed input $u(k-i)$ twice).

Guarantee

Theorem 1

As the number of data points $N \to \infty$, the solution $\widehat{\theta}$ of IV estimation converges to the true parameter vector θ_0.

Remark: This is a consistency guarantee, in the limit of infinitely many data points.

Possible extensions

- Multiple-input, multiple-output systems.
- Larger-dimension instruments Z than parameter vectors θ — with other modifications, called extended IV methods.