Lecture 6

Course Summary on System Identification

Course Outline

- Introduction and overview on system identification
- Non-recursive (off-line) identification methods
- Recursive (on-line) identification methods (III), practical aspects and applications of system identification, and summary

The System Identification Procedure

1. *Experiment design.* If possibly choose the input signal such that the data become maximally informative. Reduce the influence of noise.

2. Choose the *model structure.* Use priori knowledge and engineering intuition. Most important and most difficult step. (Do not estimate what you already know)

3. *Parameter estimation.* Determine the best model in the model structure (find optimal $\boldsymbol{\theta}$ using *e.g.*, the least squares method).

4. *Model validation.* Is the model good enough? Good is subjective, and depends on the purpose with the model.

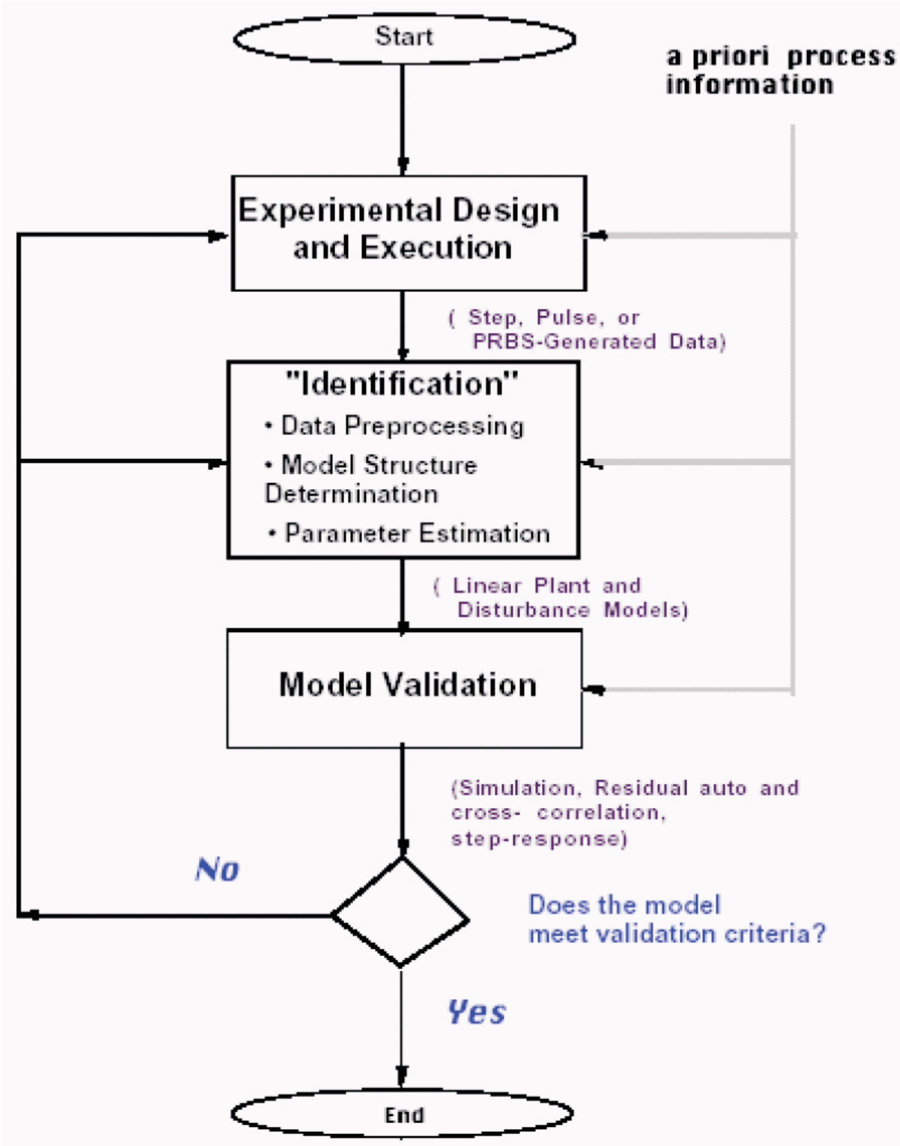

Figure 1: Procedure of System Identification

Experiment Design

- Choice of input signal.
- Choice of sampling period.
- What signals to measure, and what type of sensors to use.
- How much data is needed.
- Experimental conditions.
  - Feedback in the data?
  - Test for linearity.
  - Test for time-invariance.

Choice of Input Signal

- Signal amplitude
  - Sufficiently small to ensure that we remain in the linear region of the system.
  - Sufficiently large to ensure that we have good excitation.
- Spectral range. The input should have most of its energy in the interesting frequency regions (depends on the application).
- *Persistently exciting* of a sufficient order! $\Rightarrow$ Required to assure consistency of parametric models.
- Physical limitations.

Persistent Excitation

Define:

$$C_n = \lim_{t\to\infty} \frac{1}{t}\Phi^T\Phi = \begin{bmatrix} c(0) & c(1) & \cdots & c(n-1) \\ c(1) & c(0) & \cdots & c(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ c(n-1) & c(n-2) & \cdots & c(0) \end{bmatrix}$$

where $c(k)$ are the empirical covariances of the input. That is:

$$c(k) = \lim_{t\to\infty} \frac{1}{t}\sum_{j=1}^{t} u(j)u(j-k)$$

Definition: A signal u is called *persistently exciting* (PE) of order n if the matrix C_n is positive definite.

Examples:

- **White noise**: $u(t)$ is white noise, with zero mean and variance σ^2. Then $c(n) = \sigma^2 \delta_n$, and $C_n = \sigma^2 I_n$, which is always positive definite. Thus C_n is nonsingular for all n, and white noise signal is PE of all orders.

- **Step signal**: $u(t)$ is a step of magnitude σ, then $c(k) = \sigma^2$, and C_n is nonsingular only if $n = 1$. Then a step is PE of order 1.

- **Impulse signal**: $u(t) = 1$ for $t = 0$, and 0 otherwise. This gives $c(n) = 0$ for all n and $C_n = 0$. Therefore, this signal is not PE of any order.

Important Note: It is necessary for consistent estimation of an n-th order system that the input signal be at least persistently exciting of order $2n = n + m$, assuming $n = m$.

Determination of Model Structure

- Linear versus nonlinear, static versus dynamic, . . .
- Algorithm complexity
- Computational time and power
- Depends on the application. Simple or more sophisticated model.

Static Models

Typical examples:

- Trends and non-zero means
- Cyclic components and harmonics

Model:

$$y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}$$

where $\boldsymbol{\varphi}(t)$ is deterministic (does not depend on old values of $y(t)$).

Example: $y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}$, $\boldsymbol{\varphi}^T(t) = [1 \ t \ t^2]$.

Dynamic Models

General model:

$$y(t) = G(q^{-1})u(t) + H(q^{-1})e(t)$$
$$y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta} + v(t)$$

where $\boldsymbol{\varphi}(t)$ depends on old values of $y(t)$.

- Typical models:
  ARX: $A(q^{-1})y(t) = B(q^{-1})u(t) + \varepsilon(t)$
  ARMAX: $A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})\varepsilon(t)$
  OE: $y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + \varepsilon(t)$
  FIR: $y(t) = B(q^{-1})u(t) + \varepsilon(t)$
- The models have a certain dynamic range and are valid around a particular "working point".

Identification Method

- Nonparametric Methods
  - Transient response, frequency response, spectral analysis
  - Gives basic information about the system, with unsatisfactory accuracy, and is useful for validation
- Parametric Methods: Static and dynamic cases
  - Least squares methods, instrumental variable methods, prediction error methods
  - Good accuracy. Easy to use for *e.g.*, control
- On-line or Off-line Methods

Least Squares Methods - off-line methods

System:

$$y(t) = \varphi^T(t)\boldsymbol{\theta} + v(t), \quad t = 1, \ldots, N$$

$$\Rightarrow \qquad \boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{\theta} + \boldsymbol{v}$$

where $v(t)$ is a disturbance and $E\boldsymbol{v} = 0$, $E\boldsymbol{v}\boldsymbol{v}^T = \boldsymbol{R}$.

Estimate:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\boldsymbol{Y} = \left[\sum_{t=1}^{N} \varphi(t)\varphi^T(t)\right]^{-1}\left[\sum_{t=1}^{N} \varphi(t)y(t)\right]$$

Further Details on Least Squares Methods

For a given dynamic system

$$A(q^{-1},\boldsymbol{\theta})y(t) = B(q^{-1},\boldsymbol{\theta})u(t) + e(t)$$
$$\Rightarrow \qquad y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta} + e(t)$$

where

$$\boldsymbol{\varphi}(t) = [\,-y(t-1)\ \ldots\ -y(t-n_a)\,u(t-1)\ \ldots\ u(t-n_b)]^T$$
$$\boldsymbol{\theta} = [a_1\ \ldots\ a_{n_a}\, b_1\ \ldots\ b_{n_b}]^T$$

Problem: Find an estimate of $\boldsymbol{\theta}$ for given measurement $y(1), \boldsymbol{\varphi}(1), \ldots, y(N), \boldsymbol{\varphi}(N)$.

Solution: Introduce the *equation error*

$$\varepsilon(t) = y(t) - y_m(t) = y(t) - \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}, \quad t = 1, \ldots, N$$

or compactly

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{Y}_m = \boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{\theta}$$

Least squares method: Choose $\boldsymbol{\theta}$ such that $\varepsilon^2(t)$ is small for all t:

$$\hat{\boldsymbol{\theta}}_{LS} = \arg\min_{\boldsymbol{\theta}} V(\boldsymbol{\theta})$$

$$V(\boldsymbol{\theta}) = \frac{1}{2}\sum_{t=1}^{N} \varepsilon^2(t) = \frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{\theta})^T(\boldsymbol{Y} - \boldsymbol{\Phi}\boldsymbol{\theta})$$

Results: Assume that $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ is invertible. Then the solution of the above optimization is given by solving $\frac{\partial}{\partial \boldsymbol{\theta}} V(\boldsymbol{\theta}) = 0$, which leads to

$$\hat{\boldsymbol{\theta}}_{LS} = \left(\boldsymbol{\Phi}^T\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^T\boldsymbol{Y} = \Big[\sum_{t=1}^{N}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)\Big]^{-1}\Big[\sum_{t=1}^{N}\boldsymbol{\varphi}(t)y(t)\Big]$$

Note: The above LS algorithm is also referred as to Block/Batch LS.

Static case: Here $\boldsymbol{\varphi}(t)$ is deterministic.

- Generally consistent estimates
- For *finite* value of N we have:

$$E\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$$

$$\operatorname{cov}\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\boldsymbol{R}\boldsymbol{\Phi}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}$$

- Can be extended to include the weighted least squares and the BLUE.

Dynamic case: $\boldsymbol{\varphi}(t)$ depends on old values of $y(t)$.

- Consistent estimates if $v(t) = e(t)$ is white noise! $(Ee^2(t) = \lambda^2)$
- Asymptotically $(N \to \infty)$ it holds $(v(t) = e(t))$

$$\operatorname{cov}\hat{\boldsymbol{\theta}} = \lambda^2\left[E\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)\right]^{-1}$$

Instrumental variable methods

System:

$$y(t) = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta} + v(t), \quad t = 1, \ldots, N$$

where $v(t)$ is a disturbance with $E\boldsymbol{v} = 0$.

Estimate: Modify the least squares solution. We get:

$$\hat{\boldsymbol{\theta}} = \Big[\sum_{t=1}^{N} \boldsymbol{z}(t)\boldsymbol{\varphi}^T(t)\Big]^{-1}\Big[\sum_{t=1}^{N} \boldsymbol{z}(t)y(t)\Big]$$

where $\boldsymbol{z}(t)$ is the vector of instruments.

Comparison with the LS methods:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\boldsymbol{Y} = \Big[\sum_{t=1}^{N} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)\Big]^{-1}\Big[\sum_{t=1}^{N} \boldsymbol{\varphi}(t)y(t)\Big]$$

Results:

- Consistent estimate if:

$$
\begin{aligned}
&E\boldsymbol{z}(t)\boldsymbol{\varphi}^T(t) \quad \text{has full rank} \\
&E\boldsymbol{z}(t)v(t) = 0
\end{aligned}
$$

- The basic IV can be extended to include filtering and weighting.
- In general quite bad accuracy. Can be improved by, for instance, appropriate filtering.

Prediction error methods

Idea: Model the noise as well. General methodology applicable to a broad range of models.

The following choices have to be made:

- Choice of model structure. Example: ARMAX, OE.
- Choice of predictor $\hat{y}(t|t-1, \boldsymbol{\theta})$.
- Choice of criterion function. Example: $V(\boldsymbol{\theta}) = \frac{1}{N}\sum \varepsilon^2(t, \boldsymbol{\theta})$.

Estimate:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} V(\boldsymbol{\theta})$$

Note: "arg min" means the minimizing argument, i.e., that value of $\boldsymbol{\theta}$ which minimizes V($\boldsymbol{\theta}$).

Results:

- In general we need to perform a numerical minimization.
- Consistent estimates (if the model covers the true system).
- In general statistically efficient estimates (Gaussian noise).
- Useful also for approximations.

On-line Identification Methods

- In many cases an on-line estimate is required. e.g. adaptive signal processing, tracking time-varying parameters, fault diagnosis, etc.
- Most off-line methods can be converted into on-line methods (exactly or approximately).

On-line methods covered:

- Recursive Least Squares (RLS) Methods

RLS

Algorithm:

At time $t = 0$: Choose initial values of $\hat{\boldsymbol{\theta}}(0)$ and $\boldsymbol{P}(0)$

At each sampling instant, update $\boldsymbol{\varphi}(t)$ and compute

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{K}(t)\varepsilon(t)$$

$$\varepsilon(t) = y(t) - \boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\theta}}(t-1)$$

$$\boldsymbol{K}(t) = \boldsymbol{P}(t)\boldsymbol{\varphi}(t)$$

$$\boldsymbol{P}(t) = \left[\boldsymbol{P}(t-1) - \frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)\boldsymbol{P}(t-1)}{1 + \boldsymbol{\varphi}^T(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)}\right]$$

Question: How to obtain/derive this recursive version of LS from the block/batch LS?

Weighted RLS

Algorithm:

At time $t = 0$: Choose initial values of $\hat{\boldsymbol{\theta}}(0)$ and $\boldsymbol{P}(0)$

At each sampling instant, update $\boldsymbol{\varphi}(t)$ and compute

$$\hat{\boldsymbol{\theta}}(t) = \hat{\boldsymbol{\theta}}(t-1) + \boldsymbol{K}(t)\varepsilon(t)$$

$$\varepsilon(t) = y(t) - \boldsymbol{\varphi}^T(t)\hat{\boldsymbol{\theta}}(t-1)$$

$$\boldsymbol{K}(t) = \boldsymbol{P}(t)\boldsymbol{\varphi}(t)$$

$$\boldsymbol{P}(t) = \frac{1}{\lambda(t)}\left[\boldsymbol{P}(t-1) - \frac{\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(t)\boldsymbol{P}(t-1)}{\lambda(t) + \boldsymbol{\varphi}^T(t)\boldsymbol{P}(t-1)\boldsymbol{\varphi}(t)}\right]$$

Model Validation

A model is of no use unless it is validated!

- Compare model simulation/prediction with real data in time domain
- Compare estimated models frequency response and spectral analysis result in frequency domain
- Perform statistical tests on prediction errors

Conclusions

– System identification is a powerful technique to model dynamic systems.

– Applications in virtually all disciplines of science.

– Implemented in *e.g.*, MATLAB.

– Where to learn more: Nice textbook by Ljung, journals (Automatica) and conferences (IFAC SYSID).

Comments and Suggestions?

Questions?

You are very welcome to give me any comment or suggestion for improving the quality of the course teaching and learning!