IDENTIFICATION IN CLOSED LOOP
A powerful design tool
(theory, algorithms, applications)

better models, simpler controllers

I.D. Landau
Laboratoire d 'Automatique de Grenoble, (INPG/CNRS), France

Part 3: **Open Loop System Identification -** ***A Brief Review***

Prepared for Marie Curie Action TOK 3092

Sevilla
November 2004

OUTLINE

Open loop system identification

-Data acquisition

-Model complexity

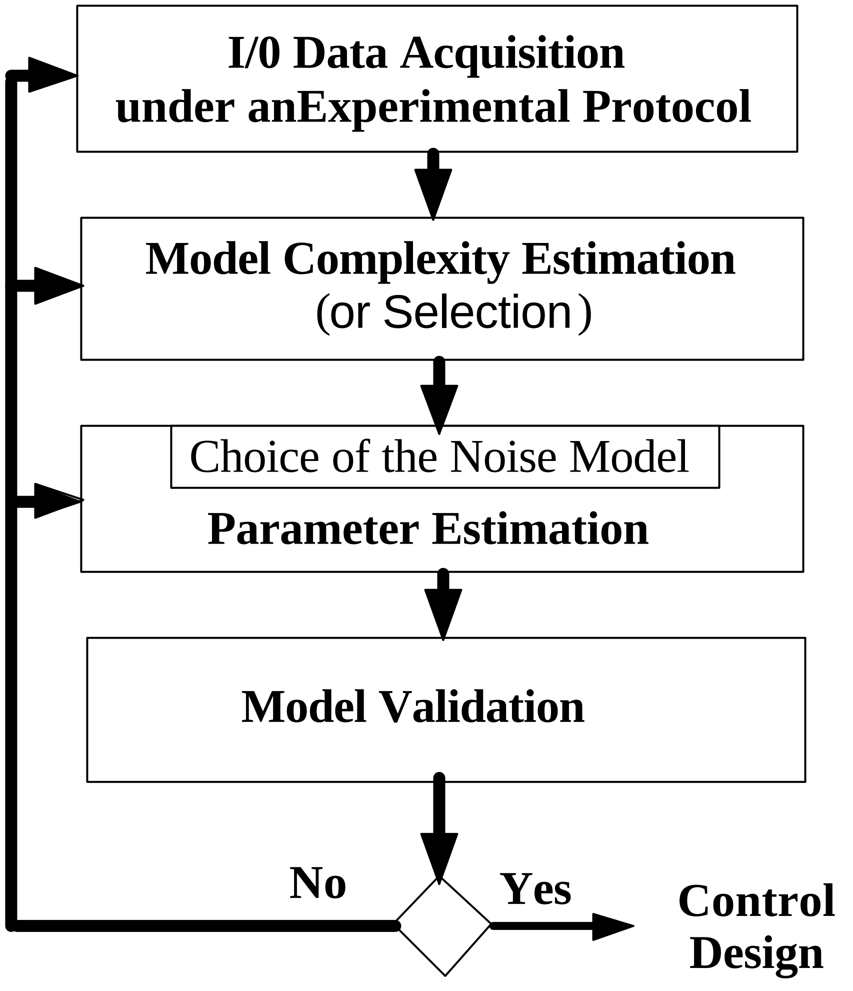
-Parameter estimation

-Validation

Objective of system identification (for control)

To extract from experimental data a dynamic model of the plant which will allow to design a controller in order to match the control specifications

System Identification Methodology



I/O Data Acqusition

Signal : *a P.R.B.S sequence*

Magnitude : *few % of the input operating point*

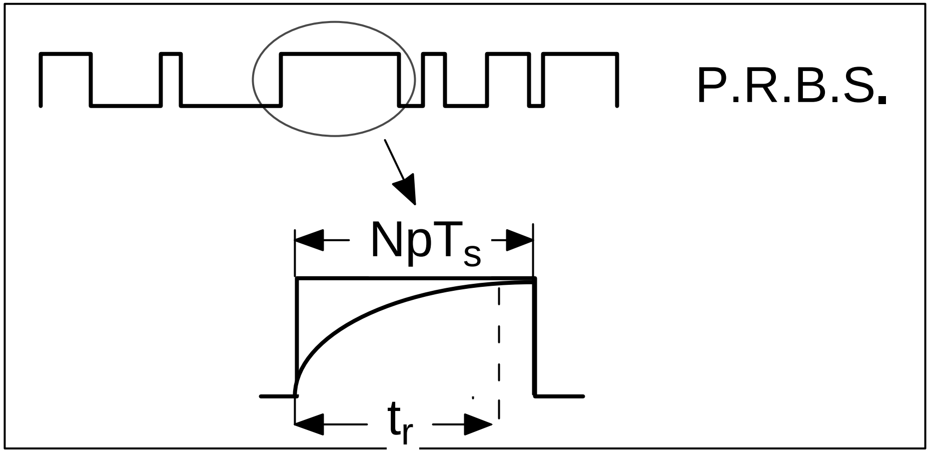
Clock frequency : $f_{clock} = (1/p) f_s$; $p = 1, 2, 3$ ($f_s = sampling\ frequency$)

Length : $(2^{N-1} - 1) pT_s$; $N = number\ of\ cells, T_s = 1/f_s$

Largest pulse : NpT_s

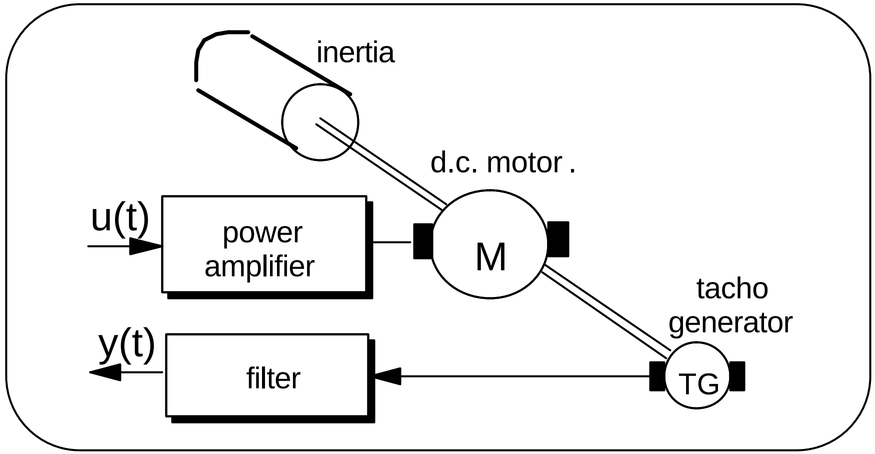
Length : < allowed duration for the experiment

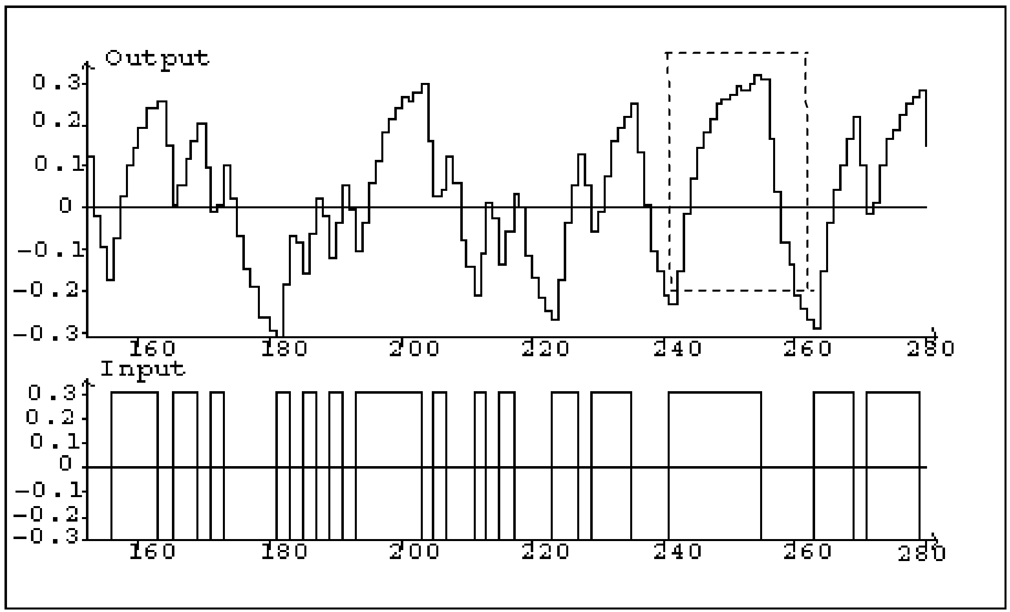
Largest pulse : $\geq t_R$ (rise time)



5

An I/O File


inertia
d.c. motor .
u(t)
power amplifier
M
tacho generator
y(t)
filter
TG


Output
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
160
180
200
220
240
260
280
Input
0.3
0.2
0.1
0
−0.1
−0.2
−0.3
160
180
200
220
240
260
280

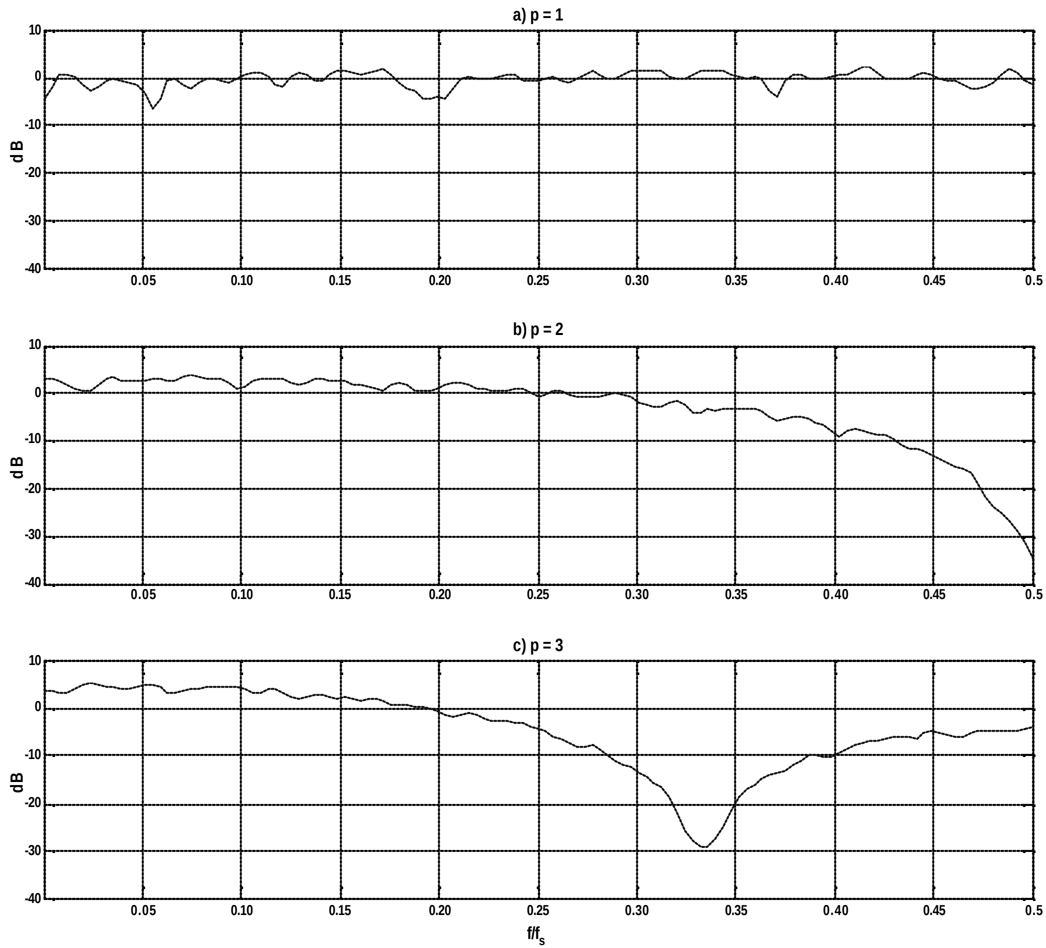
Spectral density of a P.R.B.S.

N=8

p=1
$f_{clock}=f_s$

p=2
$f_{clock}=f_s/2$

p=3
$f_{clock}=f_s/3$



Data pre-processing

The I/O data files should be centered

The use of non centered data files can cause serious errors

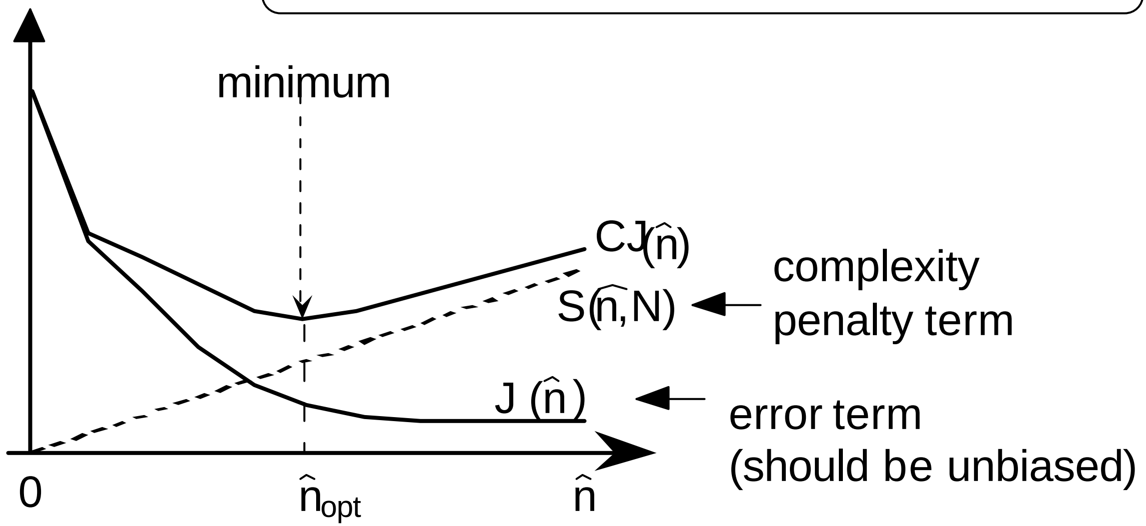
Complexity Estimation from I/O Data

Objective :

To get a good estimation of the model complexity (n_A, n_B, d) directly from noisy data

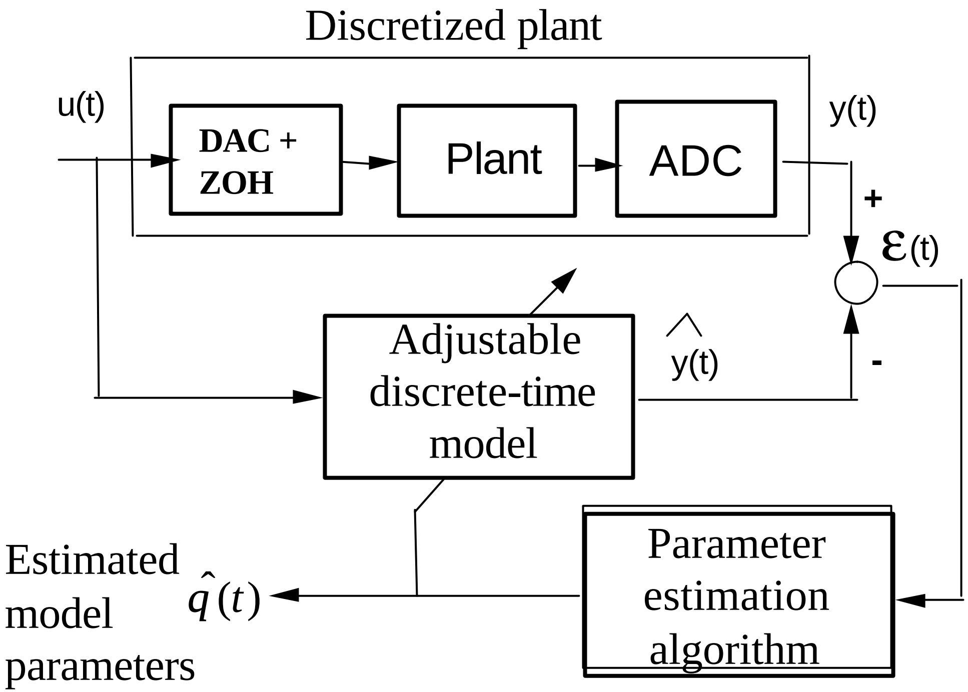
$$n = \max(n_A, n_B + d)$$

$$\hat{n}_{opt} = \min_{\hat{n}} CJ = \min_{\hat{n}} \left[J(\hat{n}) + S(\hat{n}, N) \right]$$



To get a good order estimation, J should tend to the value for noisy free data when $N \to \infty$ *(use of instrumental variables)*

Parameter Estimation

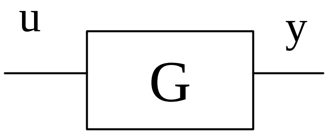


Parameter estimation alg.:
- batch (non recursive)
- recursive

It does not exists a unique algorithm providing good results in all the situations encountered in practice

Plant Model

$$G(q^{-1}) = \frac{q^{-d}B(q^{-1})}{A(q^{-1})} = \frac{q^{-d-1}B^*(q^{-1})}{A(q^{-1})}$$



$$A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_{n_A} q^{-n_A} = 1 + q^{-1} A^*(q^{-1})$$

$$B(q^{-1}) = b_1 q^{-1} + \ldots + b_{n_B} q^{-n_B} = q^{-1} B^*(q^{-1})$$

$$y(t+1) = -A^*(q^{-1})y(t) + B^*(q^{-1})u(t-d) = q^T y(t)$$

$$q^T = [a_1, \ldots a_{n_A}, b_1, \ldots, b_{n_B}]$$

$$y(t)^T = [-y(t) \ldots - y(t - n_A + 1), u(t-d) \ldots u(t - d - n_B + 1)]$$

Recursive Parameter Estimation Methods

Plant Model

$$y(t+1) = -A^*(q^{-1})y(t) + B^*(q^{-1})u(t-d) = q^T y(t)$$

$$q - parameter\ vector;\ y - measurement\ vector$$

Estimated model

$$\hat{y}^{\circ}(t+1) = \hat{q}^T(t) f(t)$$

$$\hat{q} - estimated\ parameter\ vector;\ f - observation\ vector$$

Prediction error (a priori)

$$e^0(t+1) = y(t+1) - \hat{q}^T(t)\Phi(t) = y(t+1) - \hat{y}^0(t+1)$$

Parameter adaptation algorithm (P.A.A.)

$$\hat{q}(t+1) = \hat{q}(t) + F(t+1)\Phi(t)e^0(t+1)$$

$$F^{-1}(t+1) = l_1(t)F^{-1}(t) + l_2(t)\Phi(t)\Phi^T(t)$$

$$0 < l_1(t) \le 1;\ 0 \le l_2(t) < 2$$

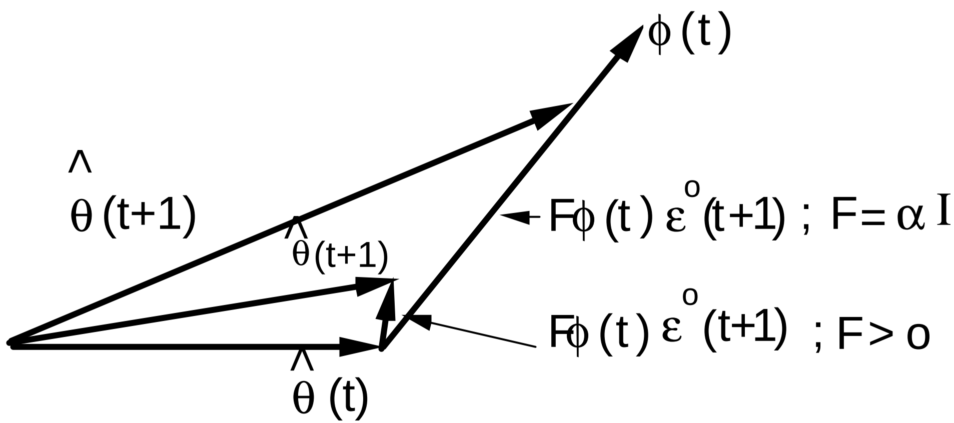
$$\Phi(t) = f[f(t)] \qquad regressor\ vector$$

Geometric interpretation of the PAA

$$\hat{q}(t+1)=\hat{q}(t)+F(t+1)f(t)e^{o}(t+1)$$

Adaptation gain

$$\begin{cases} F = aI \quad (a > 0) \\ F > 0 \quad \textit{Positive definite matrix} \end{cases}$$



One moves in the direction of the « gradient » with a deviation < 90°

Alternative form for the parameter adaptation algorithm

It can be shown (for $\Phi(t) = f(t)$) that:

$$F(t+1)\Phi(t)e^{o}(t+1) = F(t)\Phi(t)\frac{e^{o}(t+1)}{1+\Phi(t)^{T}F(t)\Phi(t)} = F(t)\Phi(t)e(t+1)$$

$$e(t+1) = y(t+1) - \overbrace{\hat{q}(t+1)f(t)}^{\hat{y}(t+1)} = \frac{e^{o}(t+1)}{1+\Phi(t)^{T}F(t)\Phi(t)}$$

a posteriori prediction error

$$\hat{q}(t+1) = \hat{q}(t) + F(t)\Phi(t)e(t+1)$$

$$F^{-1}(t+1) = l_1(t)F^{-1}(t) + l_2(t)\Phi(t)\Phi^{T}(t)$$

$$0 < l_1(t) \leq 1 ; 0 \leq l_2(t) < 2$$

$$e(t+1) = \frac{e^{o}(t+1)}{1+\Phi(t)^{T}F(t)\Phi(t)}$$

Used mainly for analysis

For $\Phi(t) \neq f(t)$ one gets the same expression, but e^{0} should be replaced sometimes by a filtered version of the a priori prediction error

Choice of the adaptation gain *F(t)*

General form:

$$F(t+1)^{-1} = l_1(t)F(t)^{-1} + l_2(t)f(t)f(t)^T$$
$$0 < l_1(t) \le 1 \ ; \ 0 \le l_2(t) < 2 \ ; \ F(0) > 0$$

$$F(t+1) = \frac{1}{l_1(t)}\left[F(t) - \frac{F(t)f(t)f(t)^T F(t)}{\frac{l_1(t)}{l_2(t)} + f(t)^T F(t)f(t)}\right]$$

A.1 Decreasing gain: $l_1(t) = l_1 = 1 \ ; \ l_2(t) = 1$ $t \nearrow F(t)^{-1} \nearrow F(t) \searrow$

Minimized criterion : $J(t) = \sum_{i=1}^{t}\left[y(i) - \hat{q}(t)^T f(i-1)\right]^2$

Identification of stationary systems (constant parameters)

Choice of the adaptation gain

Forgetting factor

A.2 Constant forgetting factor: $l_1(t) = l_1 \;\; ; \;\; 0 < l_1 < 1 \;\; ; \;\; l_2(t) = l_2 = 1$

Typical values for λ_1: $l_1 = 0.95,...,0.99$

Minimized criterion : $J(t) = \sum_{i=1}^{t} l_1^{(t-i)} \left[y(i) - \hat{q}(t)^T f(i-1) \right]^2$

Identification of slowly time varying systems

A.3 Variable forgetting factor: $l_1(t) = l_0 l_1(t-1) + 1 - l_0 \;\; ; \;\; 0 < l_0 < 1$

$l_2(t) = l_2 = 1$

Typical values: $l_1(0) = 0.95,...,0.99 \;\; ; \;\; l_0 = 0.95,...,0.99$

Minimized criterion : $J(t) = \sum_{i=1}^{t} \left[\prod_{j=1}^{t-1} l_1(j-i) \right] \left[y(i) - \hat{q}(t)^T f(i-1) \right]^2$

Since $l_1(t)$ tends towards *1* for large *i*, one forget the initial data

Recommended for the identification of stationary systems
Better performance in general than A.1

Recursive Least Squares

Plant Model

$$y(t+1) = -A^*(q^{-1})y(t) + B^*(q^{-1})u(t-d) = q^T y\ (t)$$

$$q - parameter\ vector; y\ - measurement\ vector$$

Estimated model

$$\hat{y}^{\circ}(t+1) = \hat{q}^T(t)f(t)$$

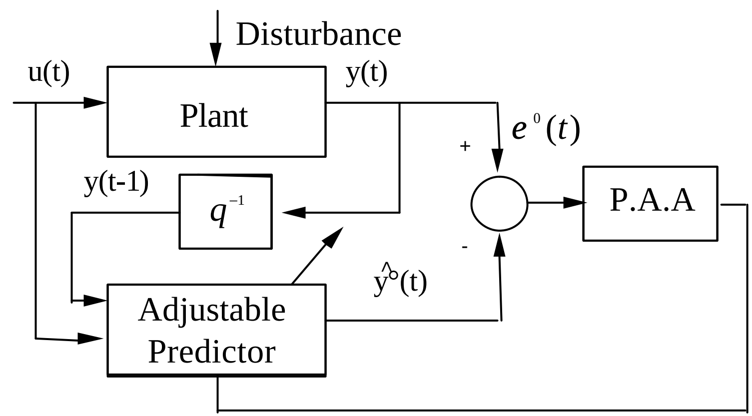
$$\hat{q} - estimated\ \ parameter\ vector; f - observation\ vector$$

$$q^T = [a_1, \ldots a_{n_A}, b_1, \ldots, b_{n_B}]$$

$$f(t)^T = y\ (t)^T = [-y(t) \ldots - y(t-n_A+1), u(t-d) \ldots u(t-d-n_B+1)]$$

Regressor vector

$$\Phi(t) = f(t) = y\ (t)$$



Effect of stochastic disturbances (noise measurement)

•Identification algorithms operates at low signal to noise ratio (in order to disturb as little as possible a plant under operation)

•This often causes an error on estimated parameters called "bias"

•The reason for the existence of many idendification algorithms is that *it does not exist an unique algorithm which gives unbiased estimates in all practical situations*

Bias in Least Squares Parameter Estimation

In the presence of measurement noise the estimation of parameters is "biased" when using least squares algorithm

Plant output in the presence of noise: $y(t+1) = q^T y(t) + w(t) = q^T f(t) + w(t)$

Bias for the least squares algorithm:

$$\hat{q}(N) = q + \left[\frac{1}{N}\sum_{t=1}^{N} f(t-1)f(t-1)^T\right]^{-1}\left[\frac{1}{N}\sum_{t=1}^{N} f(t-1)w(t)\right]$$

Condition for asymptotic unbiased estimation

$$\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N-1} f(t-1)w(t)\right] = E\{f(t-1)w(t)\} = 0 \qquad \textbf{(*)}$$

regressor (observation) vector

noise

It is necessary that f(t-1)(the regressor) and w(t) be uncorrelated

For the least squares this implies : w(t) = e(t) (white noise).

For all the other cases the estimated parameters will be biased

Unbiased estimation in the presence of noise

Suppose : $\hat{q} = q$ *and we want that the algorithm leaves unchanged this value*

$$\hat{y}(t+1|q) = q^T f(t) \longrightarrow e(t+1|q) = y(t+1) - \hat{y}(t+1|q) = w(t+1)$$

Necessary condition for unbiased estimation:

(*) ——
$$\lim_{N\to\infty}\left[\frac{1}{N}\sum_{i=1}^{N-1} f(t-1,q)e(t,q)\right] = E\{f(t-1,q)e(t,q)\} = 0$$

To eliminate the bias : $E\{f(t)e(t+1)\} = 0 \quad for \quad \hat{q} \equiv q$ necessary condition

One modifies the LS algorithm in order to obtain:

e(t+1) as a white noise for: $\hat{q} = q$

or:

uncorrelated *f(t)* and *e(t+1)* for: $\hat{q} = q$

Parameter Estimation Methods

I- *Based on the asymptotic whitening of the prediction error*
(Recursive Least Squares, Extended Least Squares, Recursive Max. Likelihood, O.E. with Extended Prediction Model)

II- *Based on the asymptotic decorrelation between the prediction error and the observation vector*
(Output Error, Instrumental Variable)

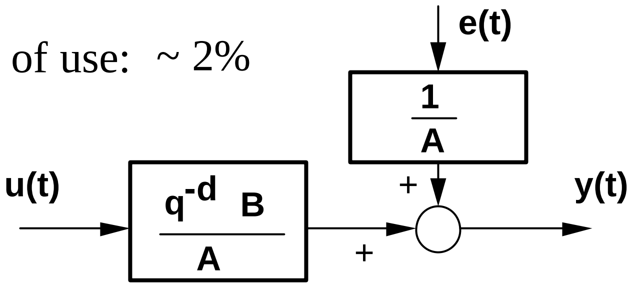
Remark:

One makes assumptions on the "noise"
and
One constructs the appropriate algorithm

«Plant + Noise » Models

$$S1: A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + e(t)$$

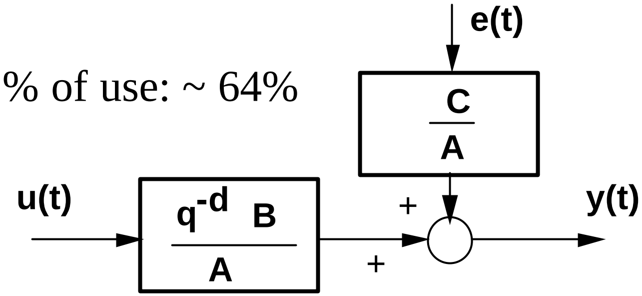
% of use: ~ 2%



(Recursive) Least Squares

$$S3: A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + C(q^{-1})e(t)$$

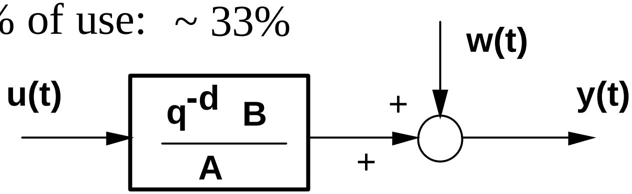
% of use: ~ 64%



Extended Least Squares
O.E. with Extended Prediction Model
(Recursive) Maximum Likelihood

$$S2: A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + A(q^{-1})w(t)$$

% of use: ~ 33%

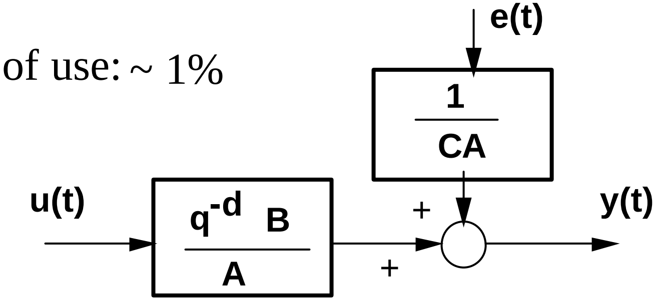


u and w are independent

Ouput Error(O.E.)
Instrumental Variable…

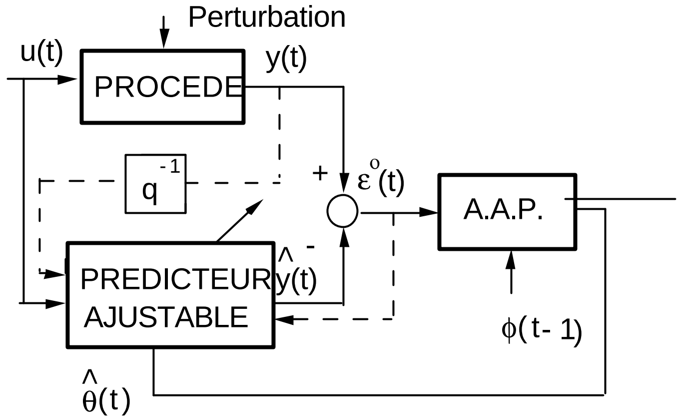
$$S4: A(q^{-1})y(t) = q^{-d}B(q^{-1})u(t) + [1/C(q^{-1})]e(t)$$

% of use: ~ 1%



Generalized Least Squares

Structure of recursive identification methods



$$\hat{q}(t+1) = \hat{q}(t) + F(t+1)\Phi(t)e^{0}(t+1)$$

$$F^{-1}(t+1) = l_1(t)F(t) + l_2\Phi(t)\Phi(t)^T$$

$$0 < l_1(t) \leq 1 \ ; \ 0 \leq l_2(t) < 2 \ ; \ F(0) > 0$$

(#)

Characteristic elements:

- predictor structure
- signals used for the observation *(f)* and regressor vectors (F)
- dimension of the vector of estimated parameters $\hat{q}$ and F
- generation of the prediction error (e)
- **all use the same parmeter adaption algorithm**

Types of identification methods:

I) *Based on the asymptotic whitening of the prediction error* (e)

II) *Based on the asymptotic decorrelation of* F and e

Extended Least Square (ELS)

Idea : *Identification of the plant model and of the disturbance (ARMAX) in order to obtain a white prediction error*

Plant + disturbance (ARMAX):

$$y(t+1) = -a_1 y(t) + b_1 u(t) + c_1 e(t) + e(t+1)$$

Optimal predictor (known parameters)

$$\hat{y}(t+1) = -a_1 y(t) + b_1 u(t) + c_1 e(t)$$

One replaces *e(t)* par *e(t)*

Predicition error (known parameters) :

$$e(t+1) = y(t+1) - \hat{y}(t+1) = e(t+1)$$

Adjustable predictor (unknown parameters):

$$\hat{y}^o(t+1) = -\hat{a}_1(t) y(t) + \hat{b}_1(t) u(t) + \hat{c}_1(t) e(t) = \hat{q}(t)^T f(t) \qquad \text{(a priori)}$$

$$\hat{q}(t)^T = \left[\hat{a}_1(t), \hat{b}_1(t), \hat{c}_1(t)\right] \qquad ; \qquad f(t)^T = \left[-y(t), u(t), e(t)\right]$$

$$\hat{y}(t+1) = -\hat{a}_1(t+1) y(t) + \hat{b}_1(t+1) u(t) + \hat{c}_1(t+1) e(t) = \hat{q}(t+1)^T f(t)^T \quad \text{(a posteriori)}$$

Regressor: $\Phi(t) = f(t)$

Extended Least Square (ELS)

Prediction error (unknown parameters)

$$e^{o}(t+1) = y(t+1) - \hat{y}^{o}(t+1) \quad ; \quad e(t+1) = y(t+1) - \hat{y}(t+1)$$

PAA: One uses (#)

Rem.: *The size of* $\hat{q}$ *and* Φ *grows with respect to the least squares*

General case :

$$\hat{q}(t)^{T} = \left[\hat{a}_1(t) \ldots \hat{a}_{n_A}, \hat{b}_1(t) \ldots \hat{b}_{n_B}(t), \hat{c}_1(t) \ldots \hat{c}_{n_C}(t)\right]$$

$$\Phi(t)^{T} = \left[-y(t) \ldots -y(t-n_A+1), u(t-d) \ldots u(t-d-n_B+1), e(t) \ldots e(t-n_C+1)\right]$$

Properties:

- *e(t) tends asymptotically towards a white noise (unbiased parameter estimation if in addition the input is persistently exciting)*
- *Sufficient convergence condition:* $\left(\frac{1}{C(z^{-1})} - \frac{l_2}{2}\right) \quad ; \quad 2 > l_2 \geq \max l_2(t)$

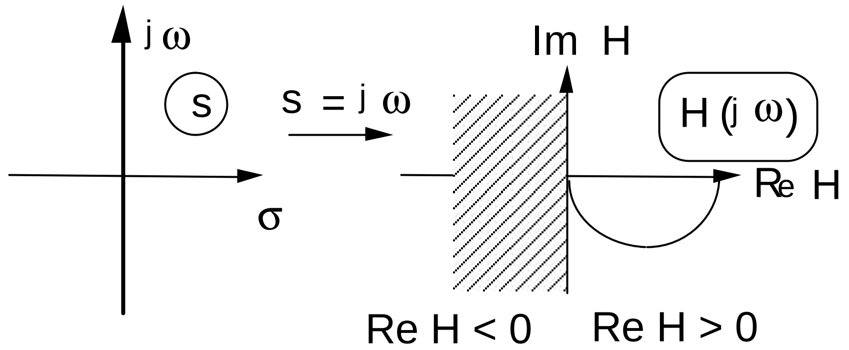
Can explain the non convergence for some noise

Strictly positive real transfer function (SPR)

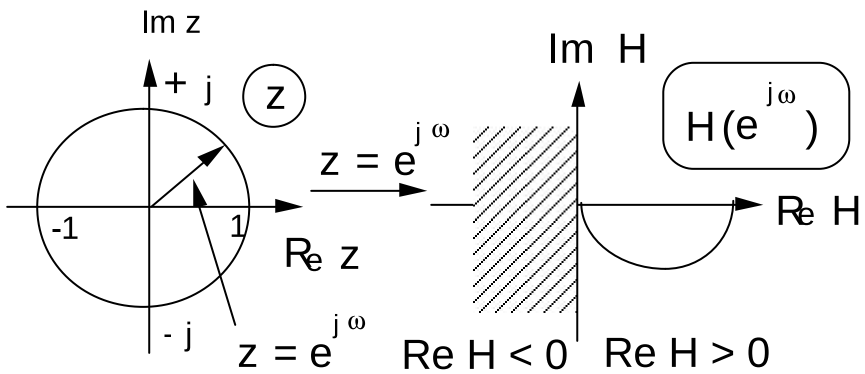
See function: ***rels.sci(.m)*** on the book web site

Strictly Positive Real Transfer Function (SRP)

Continuous time



Discrete time



- asymptotically stable
- $\operatorname{Re} H(e^{jw}) > 0 \;\; for\ all\ \left|e^{jw}\right| = 1, (0 < w < p)$ *(discrete case)*

An SPR transfer fct. introduces a phase lag less than 90° at all frequencies

Recursive output error

Plant Model

$$y(t+1) = -A^*(q^{-1})\,y(t) + B^*(q^{-1})\,u(t-d) = q^T y\,(t)$$

$$q - parameter\ vector;\ y\ - measurement\ vector$$

Estimated model

$$\hat{y}^0(t+1) = -\hat{A}^*(t,q^{-1})\,\hat{y}(t) + \hat{B}^*(t,q^{-1})\,u(t-d) = \hat{q}^T(t)\,f(t) \qquad a\ priori$$

$$\hat{y}(t+1) = \hat{q}^T(t+1)\,f(t) \qquad a\ posteriori$$

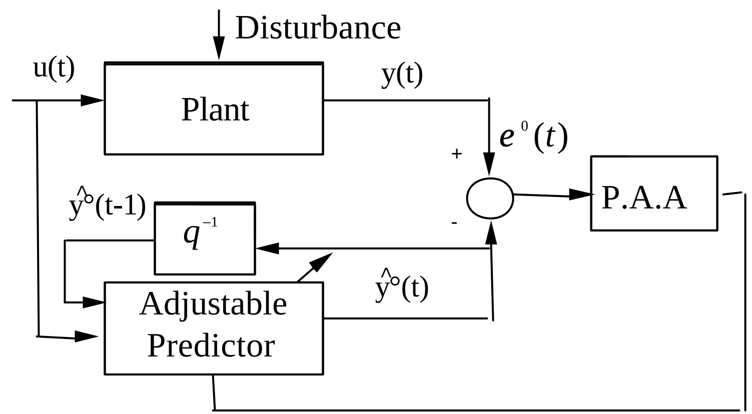
$$\hat{q} - estimated\ \ parameter\ vector;\ f - observation\ vector$$

$$q^T = \left[a_1,\ldots a_{n_A}, b_1,\ldots, b_{n_B}\right] \qquad \hat{q}^T = \left[\hat{a}_1,\ldots\hat{a}_{n_A}, \hat{b}_1,\ldots, \hat{b}_{n_B}\right]$$

$$f(t)^T = \left[-\hat{y}(t)\ldots-\hat{y}(t-n_A+1), u(t-d)\ldots u(t-d-n_B+1)\right]$$

Regressor vector

$$\Phi(t) = f(t)$$



Recursive output error

Prediction error (unknown parameters)

$$e^{o}(t+1) = y(t+1) - \hat{y}^{o}(t+1) \quad ; \quad e(t+1) = y(t+1) - \hat{y}(t+1)$$

PAA: One uses (#)

Properties:

• *Unbiased estimation of the plant parameters without identifiyng the noise model (useful for non gaussian noise)*

• *Sufficient convergence condition:*

$$\left(\frac{1}{A(z^{-1})} - \frac{l_2}{2}\right) \quad ; \quad 2 > l_2 \geq \max l_2(t)$$

Strictly positive real transfer function (SPR)

Remark:

The SPR condition can be relaxed by filtering the prediciton error or the regressor

See function ***oloe.sci (.m)*** on the web site of the book

Output error with filtered observations (OEFO)

Adjustable predictor (output error):

$$\hat{y}^{o}(t+1)=\hat{q}(t)^{T} f(t)$$

$$\hat{q}^{T}=\left[\hat{a}_1,\ldots\hat{a}_{n_A},\hat{b}_1,\ldots,\hat{b}_{n_B}\right] \qquad f(t)^{T}=\left[-\hat{y}(t)\ldots-\hat{y}(t-n_A+1),u(t-d)\ldots u(t-d-n_B+1)\right]$$

$$\hat{y}(t+1)=\hat{q}(t+1)^{T} f(t) \Rightarrow \hat{y}(t)=\hat{q}(t)^{T} f(t-1)$$

Prediction error:

$$e^{o}(t+1)=y(t+1)-\hat{y}^{o}(t+1) \quad ; \quad e(t+1)=y(t+1)-\hat{y}(t+1)$$

Filtering the observations:

Filter: $L(q^{-1})=\hat{A}(q^{-1})$ — **An estimation of the polynomial A(q-1)**

Regressor: $\Phi(t)=f_f(t)=\dfrac{1}{\hat{A}(q^{-1})} f(t)$

PAA: One uses (#) with : $\Phi(t)=f_f(t)$

- *Sufficient convergence condition :* $\left(\dfrac{\hat{A}(z^{-1})}{A(z^{-1})}-\dfrac{l_2}{2}\right) \quad ; \quad 2>l_2 \geq \max l_2(t)$

Strictly positive real transfer function (SPR)

See function ***foloe.sci(.m)*** on the book web site

Output error with adaptive filtered observations (OEAFO)

Uses an adaptive filter on the observations instead of a fixed filter
(takes advantage of the improvement of the estimation of A as times goes)

Filtering the observations:

Filtre: $L(t, q^{-1}) = \hat{A}(t, q^{-1})$ — **Estimation of polynomial *A(q⁻¹)* provided by the algorithm itself**

$$\Phi(t) = f_f(t) = \frac{1}{\hat{A}(t, q^{-1})} f(t)$$

Initialization:

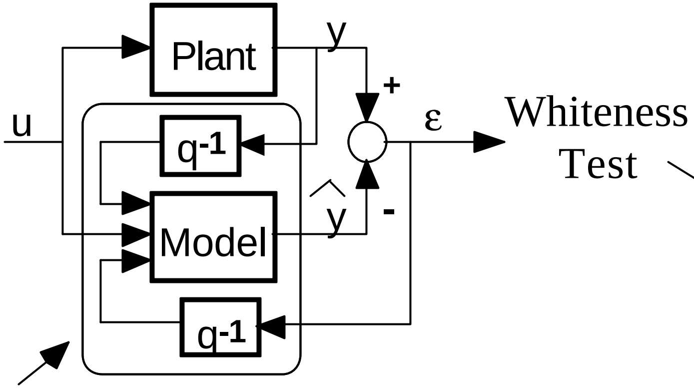
$\hat{A}(0, q^{-1}) = \hat{A}_0(q^{-1})$ or: $\hat{A}(0, q^{-1}) = 1$ (simpler and more efficient)

Remove in most of the case the problems related to the SPR condition

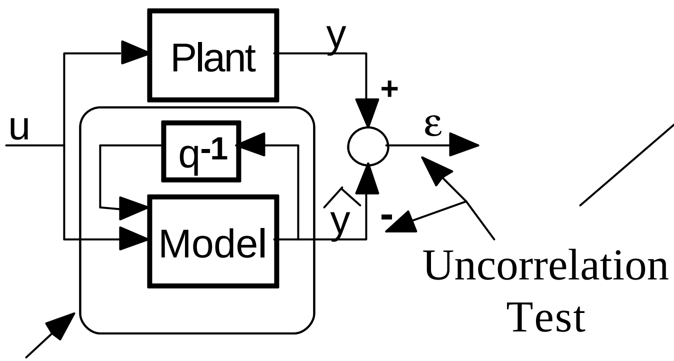
See function: ***afoloe.sci(.m)*** on the book web site

Validation of Identified Models

Statistical Validation



ARMAX (ARARX) predictor

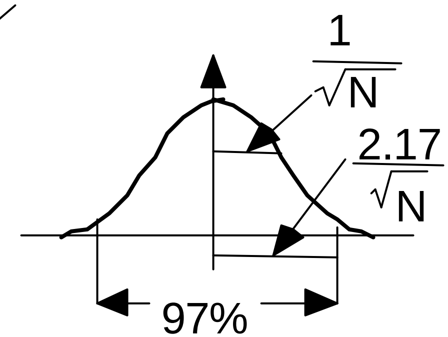


Output Error Predictor

$$|RN(i)| \leq \frac{2.17}{\sqrt{N}}\ ;\ i \geq 1$$

Normalized autocorrelations or crosscorelations

number of data



$$N = 256 \rightarrow |RN(i)| \leq 0.136$$

pratical value : $|RN(i)| \leq 0.15$

«Whiteness » test

{e(t)} : centered sequence of residual prediction errors

One computes:

$$R(0) = \frac{1}{N}\sum_{t=1}^{N} e^2(t) \quad ; \quad RN(0) = \frac{R(0)}{R(0)} = 1$$

$$R(i) = \frac{1}{N}\sum_{t=1}^{N} e(t)e(t-i) \quad ; \quad RN(i) = \frac{R(i)}{R(0)} \quad ; \quad i = 1,2,3\ldots i_{\max} \ ;$$

Theoretical values: $RN(i) = 0;\ i = 1, 2\ldots i_{max}$

Real situation:
- Finite number of data
- Residual structural errors (orders, nonlinearities, noise)
- Objective: to obtain « good » simple models

Validation criterion (*N* = number of data):

$$\left|RN(i)\right| \le \frac{2.17}{\sqrt{N}} \quad ; \quad i \ge 1$$

or: $\left|RN(i)\right| \le 0.15\ ;\ i = 1,\ldots,i_{\max}$

« Uncorrelation » test

$\{e(t)\}, \{\hat{y}(t)\}$: centered sequences of residual prediction errors and predictions

Output error predictor: $\hat{A}(q^{-1})\hat{y}(t) = q^{-d}\hat{B}(q^{-1})u(t)$

One computes:

$$R(i) = \frac{1}{N}\sum_{t=1}^{N} e(t)\hat{y}(t-i) \quad ; \quad i = 0,1,2,\ldots,i_{\max} \qquad ; \qquad i_{\max} = \max(n_A, n_B + d)$$

$$RN(i) = \frac{R(i)}{\left[\left(\frac{1}{N}\sum_{t=1}^{N}\hat{y}^2(t)\right)\left(\frac{1}{N}\sum_{t=1}^{N}e^2(t)\right)\right]^{1/2}} \quad ; \quad i = 0,1,2,\ldots,i_{\max}$$

Remark: $RN(0) \neq 1$

Theoretical values: *RN(i) =0; i = 1, 2…*i_{max}

Real situation:

- Finite number of data
- Residual structural errors (orders, nonlinearities, noise)
- Objective: to obtain « good » simple models

Validation criterion (N = number of data):

$$|RN(i)| \leq \frac{2.17}{\sqrt{N}} \quad ; \quad i \geq 1$$

or: $|RN(i)| \leq 0.15\,;\ i = 1,\ldots,i_{\max}$

Open loop system identification - references

More details can be found in :

I.D. Landau *System identification and control design*, Prentice Hall,NJ, 1990

I.D. Landau: *Commande des systèmes – conception, identification, mise en œuvre*, Hermes, 2002, Paris. Chapters 5,6 (translation available) and 7

and

http://landau-bookic.lag.ensieg.inpg.fr

Free routines (matlab, scilab) and slides can be downloaded

General references:

T. Söderstrom, P. Stoica : *System identification*, Prentice Hall, UK, 1989

L. Ljung : *System identification – Theory for the user*, Prentice Hall, NJ, 2nd edition, 2002