Model quality in nonlinear SM identification

Mario Milanese e Carlo Novara

Dipartimento di Automatica e Informatica
Politecnico di Torino

42nd IEEE CDC

Maui, December 9-12, 2003
Outline

• Introduction
• Nonlinear SM identification
• Estimation of the simulation error
• Example
• Conclusions
Introduction

• Consider a nonlinear dynamic system of the form:

\[ y_{t+1} = f_0(w_t) \]
\[ w_t = [y_t...y_{t-n+1} u_{t-1}u_{t-n+1}]^T \in W \subset \mathbb{R}^m \]

• \( f_0 \) is not known.

• A set of noise corrupted measurements \( \tilde{y}_t \) and \( \tilde{w}_t \), of \( y_t \) and \( w_t \), \( t = 0, 1, 2, ..., T \) is available.

Problem:

Find estimate \( \hat{f} \) of \( f_0 \)
giving “small” simulation error
for any future input sequence
Introduction

- Most of identification methods in the literature assume:
  \[ f_o \in K \equiv \{ f_p(\varphi), p \in \mathbb{R}^r, \varphi \in \mathbb{R}^m \} \]
  Measured data are used to derive an estimate \( \hat{p} \) of \( p \).

- Estimate \( \hat{p} \) of \( p \) is usually obtained by means of a Prediction Error (PE) method:
  \[
  \hat{p} = \arg \min_p V(p, \Phi_T) \\
  V(p, \Phi_T) = \sum_{t=0}^{T-1} |\tilde{y}_{t+1} - f_p(\varphi_t)|^l
  \]
  where \( \varphi_t \) is a regression vector and \( \Phi_T = [\varphi_0, \varphi_1, \ldots, \varphi_T] \).

- Widely used are the following choices for the regressor \( \varphi_t \):
  \[
  \varphi_t = \tilde{w}_t = [\tilde{y}_t \ldots \tilde{y}_{t-n+1} \tilde{u}_t \ldots \tilde{u}_{t-n+1+1}]^T \implies NARX \\
  \varphi_t = \hat{w}_t = [f_p(\tilde{w}_{t-1}) \ldots f_p(\tilde{w}_{t-n}) \tilde{u}_t \ldots \tilde{u}_{t-n+1+1}]^T \implies NOE
  \]
Introduction

Problems:

- Models giving lower cost functions (prediction errors) do not necessarily give lower simulation errors on future inputs.
- Even boundedness of the simulation error is not guaranteed.

More relevant for NARX than for NOE models.

Nonlinear SM identification

- Let $\tilde{y}_t$ and $\tilde{w}_t$ noise corrupted data generated by the system $y_{t+1} = f_o(w_t)$. Then:
  $$\tilde{y}_{t+1} = f_o(\tilde{w}_t) + e_t, \ t = 0, 1, .., T - 1$$

- Let $\theta \in \mathbb{R}^m$ be a linear approximation of $f_o$, i.e. $f_o(w) \approx \theta^T w$, and let:
  $$f_\Delta(w) = f_o(w) - \theta^T w$$
  $f_\Delta(w)$ is called residue function.

Assumptions on $f_\Delta(w)$:
$$f_\Delta \in K^L \doteq \{ g \in C^1(W), \|g'(w)\| \leq \gamma, \forall w \in W \}$$
$g'(w)$: gradient of $g(w)$, $\|w\|$: euclidean norm.

Assumptions on noise:
$$|e_t| \leq \varepsilon_t, \ t = 0, 1, ..., T$$
Nonlinear SM identification

• **Feasible Systems Set:**

\[
FSST_T \doteq \{ f : f(w) = \theta^T w + g(w), \ g \in K^L, \\
|\tilde{y}_{t+1} - f(\tilde{w}_t)| \leq \varepsilon_t, \ t = 0, 1, \ldots, T - 1 \}
\]

• **Identification error of estimate \( \hat{f} \):**

\[
E(\hat{f}) \doteq \sup_{f \in FSST} \| f - \hat{f} \|_p
\]

• **Optimal estimate:**

\[
E(f^*) \doteq \inf_{f} E(f) = \inf_{\hat{f}} \sup_{f \in FSST} \| f - \hat{f} \|_p = r_I
\]

\( r_I \): (local) radius of information, i.e. minimal identification error that can be guaranteed by any estimate based on the available information up to time \( T \).

\[
\| f \|_p \doteq \left( \int_W |f(w)|^p \, dw \right)^{1/p}, \ p < \infty, \ \| f \|_\infty \doteq \text{ess-sup}_{w \in W} |f(w)|, \ W: \text{bounded subset of } \mathbb{R}^m.
\]
Nonlinear SM identification

- Define:

\[ f_c(w) \triangleq \theta^T w + \frac{1}{2} \left[ f_\Delta(w) + \overline{f}_\Delta(w) \right] \]

\[ \overline{f}_\Delta(w) \triangleq \min_{t=0,\ldots,T-1} (h_t + \gamma \| w - \tilde{w}_t \|) \]

\[ \underline{f}_\Delta(w) \triangleq \max_{t=0,\ldots,T-1} (h_t - \gamma \| w - \tilde{w}_t \|) \]

\[ \overline{h}_t \triangleq \tilde{y}_{t+1} - \theta^T \tilde{w}_t + \varepsilon_t \quad h_k \triangleq \tilde{y}_{t+1} - \theta^T \tilde{w}_t - \varepsilon_t \]

**Theorem 1** For any \( L_p(W) \) norm, with \( p \in [1, \infty] \):

i) The estimate \( f_c \) is optimal

ii) \( E(f_c) = \frac{1}{2} \left\| \overline{f}_\Delta - \underline{f}_\Delta \right\|_p = r_I = \inf_f E(f) \)
Estimation of the simulation error

• Consider the nonlinear system:

\[ y_{t+1} = f_o(x_t, v_t) \]
\[ x_t = [y_t \ldots y_{t-n+1}]^T \]
\[ v_t = [u_1^t \ldots u_{t-n+1}^1 \ldots u_q^t \ldots u_{t-nq+1}^q]^T \]

• For given initial condition \( x_0 \in X \) and input sequence \( v = [v_0, v_1, v_2, \ldots] \) the sequence:

\[ y_t(f, x_0, v), t = 0, 1, 2, \ldots \]

is called solution of the system, corresponding to initial condition \( x_0 \) and input \( v \).

• The simulation error at time \( t \) of model \( y_{t+1} = f_c(x_t, v_t) \) is:

\[ SE_t = |y_t(f_o, x_0, v) - y_t(f_c, \tilde{x}_0, v)| \]

Being \( f_o \) and \( x_0 \) not known, \( SE_t \) cannot be exactly evaluated, and a bound on it is looked for.
Estimation of the simulation error

• Let:

\[
\Theta = \begin{bmatrix}
\theta_1 & \theta_2 & \cdots & \theta_{n-1} & \theta_n \\
1 & 0 & \cdots & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 1 & 0 \\
\end{bmatrix} \in \mathbb{R}^{n \times n}
\]

Note:

\[|\lambda_i(\Theta)| < 1\]

\[\downarrow\]

\[\|\Theta^t\| \leq L\rho^t, \ \forall t\]

for some \(L > 0\) and \(\max_i |\lambda_i(\Theta)| \leq \rho < 1\)
Estimation of the simulation error

Theorem 2 Assume that:

i) \(|\lambda_i(\Theta)| < 1\)

ii) \(\gamma < \frac{1-\rho}{L}\)

Then, for all initial conditions \(x_o\) and inputs \(v\) giving solutions for \(f_o\) such that \((x_t, v_t) \in W\ \forall t\), a constant \(K \in [0, \infty)\) exists such that the simulation error \(SE_t\) is bounded as:

\[
SE_t \leq K r_I = \frac{K}{2} \left\| \bar{f}_\Delta - f_\Delta \right\|_\infty \quad \forall t
\]

Note:

\(|\lambda_i(\Theta)| < 1\)

\(\Downarrow\)

The linear regression model \(y_{t+1} = \theta^T w_t\) is asymptotically stable.
Example

- A set of 6000 data has been generated from the nonlinear system:
  \[ y_{t+1} = 1.8y_t - 0.82y_{t-1} + 0.0024\sin(y_{t-1}) + 0.047\tanh(3u_t) \]

![Figure 1: Nonlinear mass-spring-damper system.](image)

- A random input of amplitude ≤ 1 has been used.
- The output data of the estimation set have been corrupted by a uniform random additive noise of amplitude ≤ 0.025.
Example

Figure 2: Estimation data set (bold line) and validation data set (dashed line).

- **Estimation set**: the first 5000 data, called estimation set, used for model identification.

- **Validation set**: the remaining 1000 data, used for model testing.
Example

- Regressions of the following form have been considered for model identification:

\[
\begin{align*}
y_{t+1} &= f(w_t) \\
w_t &= [y_t \ y_{t-1} \ u_t]^T
\end{align*}
\]

**Linear Output Error model OE:**

\[
f(w) = \theta^T w
\]

where \( \theta = [1.8 \ -0.81 \ 0.06]^T \) has been estimated by means of the Matlab Systems Identification Toolbox using the output error estimation method.
Example

Nonlinear Set Membership model NSM:

\[ f_c(w) = \theta^T w + \frac{1}{2} \left[ f_\Delta(w) + \overline{f_\Delta}(w) \right] \]

\[ L = 19.8 \quad \rho = 0.952 \quad \gamma = 0.0024 \quad \varepsilon = 0.08 \]

Figure 3: Validation curve for model NSM.

Figure 4: \|\Theta^t\| (bold line) and \(L\rho^t\) (thin line) sequences.
Example

Neural Network models $\text{NN}_{\text{narx}}$ and $\text{NN}_{\text{noe}}$:

$$f(w) = \sum_{i=1}^{r} \alpha_i \sigma \left( \beta_i^T w - \lambda_i \right) + \zeta$$

- Several NARX and NOE models with different values of $r$ (from $r = 3$ to $r = 16$) have been trained using the Matlab Neural Networks Toolbox.
- The NARX model with $r = 8$ showing the best simulation performances, has been taken for model $\text{NN}_{\text{narx}}$.
- All the NOE identified models got stuck on (possibly) local minima during the training phase, providing bad simulation performances.
- The best result has been obtained by using as starting point the parameters of the $\text{NN}_{\text{narx}}$ model. This NOE model, showing a slight improvement in simulation performances with respect to the $\text{NN}_{\text{narx}}$ one, has been taken for model $\text{NN}_{\text{noe}}$. 
Example

- In table 1 the root mean square errors obtained by the identified models on the validation data set are reported.

**RMSE\textsubscript{P}:** one-step ahead prediction error  
**RMSE\textsubscript{S}:** simulation error

<table>
<thead>
<tr>
<th>Model</th>
<th>NSM</th>
<th>OE</th>
<th>NN\textsubscript{narx}</th>
<th>NN\textsubscript{noe}</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE\textsubscript{P}</td>
<td>0.005</td>
<td>0.011</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>RMSE\textsubscript{S}</td>
<td>0.091</td>
<td>0.267</td>
<td>0.299</td>
<td>0.262</td>
</tr>
</tbody>
</table>

Table 1. One-step ahead prediction and simulation errors.

Figure 5: Validation set: data (bold line), NSM simulation (thin line) and NNnoe simulation (dashed line).
Conclusions

• The quality of identified models is related to the accuracy in simulating the system behavior for future inputs not used in the identification.

• Models identified by classical methods minimizing the prediction error, do not necessarily give “good” simulation error on future inputs and even boundedness of this error is not guaranteed.

• Using a Set Membership approach, under suitable conditions on the bounding constants $\gamma$ and $\varepsilon$ defining the SM assumptions, the simulation error can be bounded as a function of the radius of information $r_I$ that goes to zero as $r_I$ decreases to zero.