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:

$$y_{t+1} = f_o(w_t)$$
$$w_t = [y_t \ldots y_{t-n+1} \; u_t \ldots u_{t-n_1+1}]^T \in W \subset \mathbb{R}^m$$

- f_o is **not known**.
- A set of noise corrupted measurements $\widetilde{y}_t$ and $\widetilde{w}_t$, of y_t and w_t, $t = 0, 1, 2, \ldots, T$ is available.

Problem:

Find estimate $\widehat{f}$ of f_o
giving "small" simulation error
for any future input sequence

Introduction

- Most of identification methods in the literature assume:

$$f_o \in K \doteq \{f_p(\varphi), p \in \mathbb{R}^r, \varphi \in \mathbb{R}^m\}$$

Measured data are used to derive an estimate $\widehat{p}$ of p.

- Estimate $\widehat{p}$ of p is usually obtained by means of a Prediction Error (PE) method:

$$\widehat{p} = \arg\min_p V(p, \Phi_T)$$
$$V(p, \Phi_T) = \sum_{t=0}^{T-1} \left|\widetilde{y}_{t+1} - f_p(\varphi_t)\right|^l$$

where φ_t is a regression vector and $\Phi_T = [\varphi_0, \varphi_1, \dots, \varphi_T]$.

- Widely used are the following choices for the regressor φ_t:

$$\varphi_t = \widetilde{w}_t = [\widetilde{y}_t \dots \widetilde{y}_{t-n+1}\ \widetilde{u}_t \dots \widetilde{u}_{t-n_1+1}]^T \quad \Longrightarrow \quad NARX$$

$$\varphi_t = \widehat{w}_t = [f_p(\widehat{w}_{t-1}) \dots f_p(\widehat{w}_{t-n})\ \widetilde{u}_t \dots \widetilde{u}_{t-n_1+1}]^T \quad \Longrightarrow \quad NOE$$

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.

$$\Downarrow$$

More relevant for NARX than for NOE models.

- Using the Nonlinear SM identification method (*M. Milanese and C. Novara, "Optimality in SM Identification of Nonlinear Systems", SYSID 2003*) it is possible to derive conditions for boundedness of simulation error.

Nonlinear SM identification

- Let $\widetilde{y}_t$ and $\widetilde{w}_t$ noise corrupted data generated by the system $y_{t+1} = f_o\left(w_t\right)$. Then:

$$\tilde{y}_{t+1} = f_o\left(\widetilde{w}_t\right) + e_t,\ t = 0, 1, .., T-1$$

- Let $\theta \in \mathbb{R}^m$ be a linear approximation of f_o, i.e. $f_o\left(w\right) \approx \theta^T w$, and let:

$$f_\Delta\left(w\right) \doteq f_o\left(w\right) - \theta^T w$$

$f_\Delta\left(w\right)$ is called *residue function*.

Assumptions on $f_\Delta\left(w\right)$:

$$f_\Delta \in K^L \doteq \{g \in C^1(W), \|g'\left(w\right)\| \leq \gamma, \forall w \in W\}$$

$g'(w)$: 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*:

$$FSS_T \doteq \{f : f(w) = \theta^T w + g(w),\ g \in K^L,\\ |\tilde{y}_{t+1} - f(\widetilde{w}_t)| \le \varepsilon_t,\ t = 0, 1, ..., T-1\}$$

■

- *Identification error of estimate $\hat{f}$:*

$$E(\hat{f}) \doteq \sup_{f \in FSS_T} \left\| f - \hat{f} \right\|_p$$

■

- *Optimal estimate:*

$$E(f^*) \doteq \inf_{f} E(f) = \inf_{\hat{f}} \sup_{f \in FSS_T} \left\| f - \hat{f} \right\|_p = r_I$$

■

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

Nonlinear SM identification

- Define:

$$f_c(w) \doteq \theta^T w + \frac{1}{2}\left[\underline{f}_\Delta(w) + \overline{f}_\Delta(w)\right]$$
$$\overline{f}_\Delta(w) \doteq \min_{t=0,\dots,T-1} \left(\overline{h}_t + \gamma \left\| w - \widetilde{w}_t \right\|\right)$$
$$\underline{f}_\Delta(w) \doteq \max_{t=0,\dots,T-1} \left(\underline{h}_t - \gamma \left\| w - \widetilde{w}_t \right\|\right)$$
$$\overline{h}_t \doteq \widetilde{y}_{t+1} - \theta^T \widetilde{w}_t + \varepsilon_t \quad \underline{h}_k \doteq \widetilde{y}_{t+1} - \theta^T \widetilde{w}_t - \varepsilon_t$$

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:

$$y_{t+1} = f_o(x_t, v_t)$$
$$x_t = [y_t \ldots y_{t-n+1}]^T \quad v_t = [u_t^1 \ldots u_{t-n_1+1}^1 \ldots u_t^q \ldots u_{t-n_q+1}^q]^T$$

- 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:

$$SE_t \doteq |y_t(f_o, x_0, v) - y_t(f_c, \widetilde{x}_0, v)|$$

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:

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

Note:

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

$$\Downarrow$$

$$\left\|\Theta^t\right\| \leq L\rho^t, \ \forall t$$

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:*

$$SE_t \leq K r_I = \frac{K}{2} \left\| \overline{f}_\Delta - \underline{f}_\Delta \right\|_\infty \quad \forall t$$

■

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:

$$y_{t+1} = 1.8y_t - 0.82y_{t-1} + 0.0024\sin(y_{t-1}) + 0.047\tanh(3u_t)$$

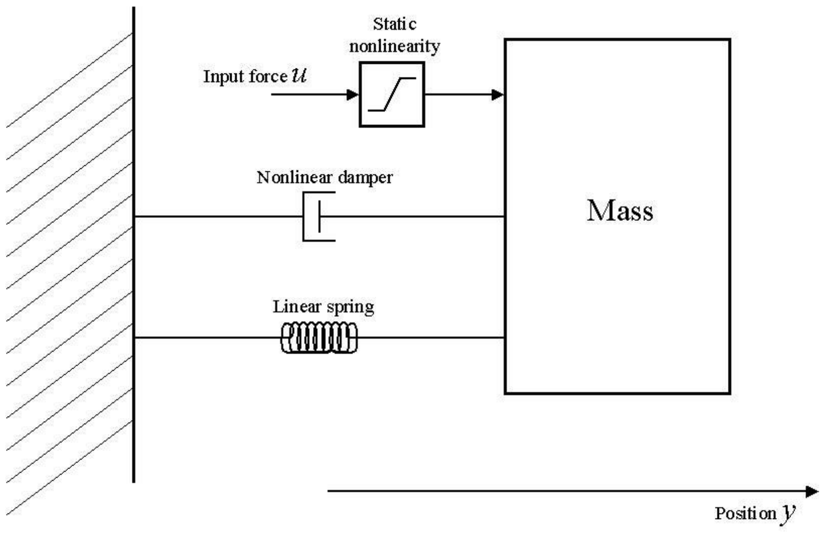

Figure 1: Nonlinear mass-spring-damper system.

- 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

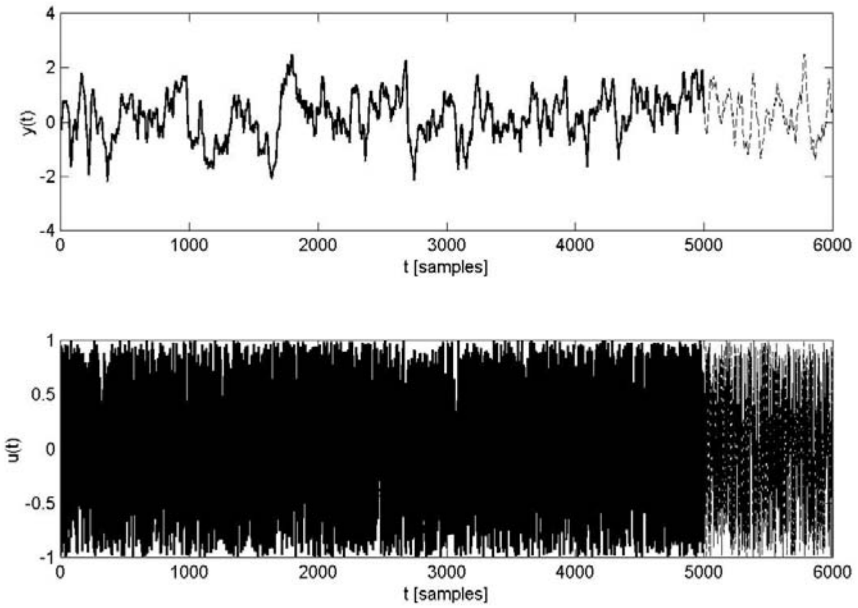

Figure 2: Estimation data set (bold line) and validation data set (dashed line).

- **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:

$$y_{t+1} = f(w_t)$$
$$w_t = [y_t\ y_{t-1}\ u_t]^T$$

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:

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

$$L = 19.8 \qquad \rho = 0.952 \qquad \gamma = 0.0024 \qquad \varepsilon = 0.08$$

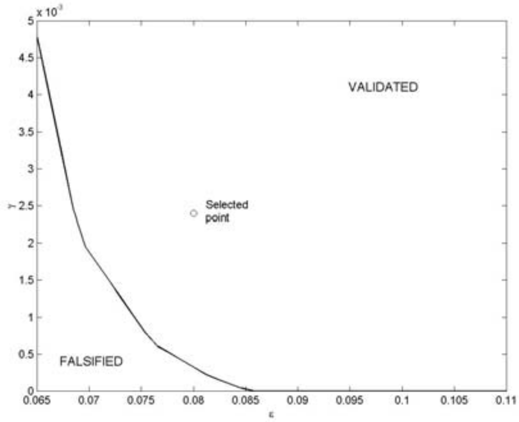

Figure 3: Validation curve for model NSM.

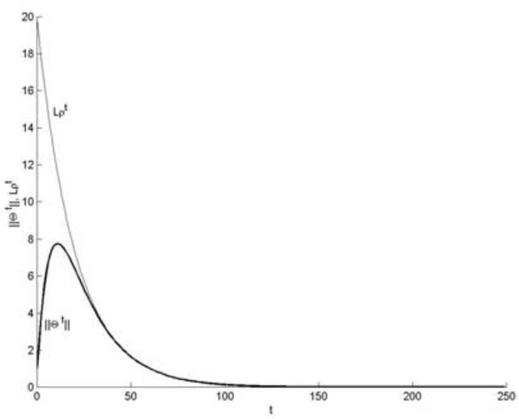

Figure 4: $\|\Theta^t\|$ (bold line) and $L\rho^t$ (thin line) sequences.

Example

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

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

- 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 $\mathrm{NN}_{\mathrm{narx}}$.
- 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 $\mathrm{NN}_{\mathrm{narx}}$ model. This NOE model, showing a slight improvement in simulation performances with respect to the $\mathrm{NN}_{\mathrm{narx}}$ one, has been taken for model $\mathrm{NN}_{\mathrm{noe}}$.

Example

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

RMSE_P: one-step ahead prediction error
RMSE_S: simulation error

Model	NSM	OE	NN_narx	NN_noe
RMSE_P	0.005	0.011	0.008	0.009
RMSE_S	0.091	0.267	0.299	0.262

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

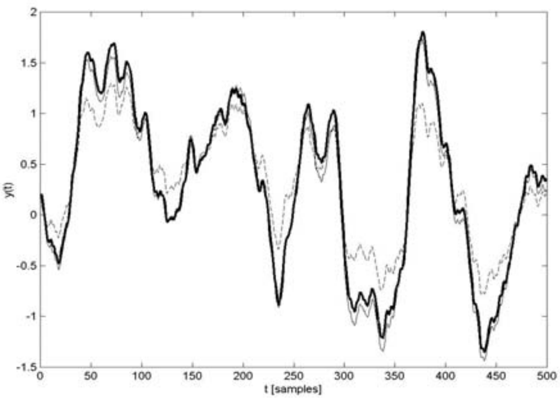

Figure 5: Validation set: data (bold line), NSM simulation (thin line) and NNnoe simulation (dashed line).

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 necessary 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 γ and ε 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.