Nonlinear systems identification

Mario Milanese, Carlo Novara

Dipartimento di Automatica e Informatica
Politecnico di Torino
The nonlinear system ID problem

Data are generated by the nonlinear system $f^o$:

$$y^{t+1} = f^o(w^t)$$

$$w^t = [y^t \ldots y^{t-n_y} u^t \ldots u^{t-n_u}]$$

$u^t$: known variables

The system $f^o$ is unknown, but a finite number of noise corrupted measurements of $y^t$, $w^t$ are available:

$$\tilde{y}^{t+1} = f^o(\tilde{w}^t) + d^t, \quad t = 1, \ldots, N$$

$d^t$ accounts for errors in data $\tilde{y}^t$, $\tilde{w}^t$

Identification problem: find an estimate $\hat{f} \cong f^o$
The nonlinear system ID problem

Related problems:

- For a given estimate \( \hat{f} \approx f^o \)
  
  evaluate the identification error \( \left\| f^o - \hat{f} \right\| \)

- Find an estimate \( \hat{f} \approx f^o \)
  
  “minimizing” the identification error

The identification error cannot be exactly evaluated since \( f^o \) and \( d^i \) are not known

Need of prior assumptions on \( f^o \) and \( d^i \) for deriving finite bounds on the identification error
**Parametric approach**

- **Typical assumptions in literature:**
  - on system: \( f^o \in \Psi(\theta) = \left\{ f(w, \theta) = \sum_{i=1}^{r} \alpha_i \sigma(w, \beta) \right\} \)
  - on noise: iid stochastic

- **Functional form of \( f^o \):**
  - derived from physical laws
  - \( \sigma_i \): "basis" function (polynomial, sigmoid, ..).

- **Parameters \( \theta \) are estimated by means of the Prediction Error (PE) method.**
**Parametric approach**

- **Predictor:**
  \[
  \hat{y}^{t+1} = f(w^t, \theta) = \sum_{i=1}^{r} \alpha_i \sigma(w^t, \beta_i)
  \]

- **Given** \( N \) noise-corrupted measurements of \( y^t, w^t \):

  \[
  y^2 = f(w^1, \theta) + \varepsilon^2 \\
  y^3 = f(w^2, \theta) + \varepsilon^3 \\
  \vdots \\
  y^{N+1} = f(w^N, \theta) + \varepsilon^{N+1}
  \]

  \[
  Y = F(\theta) + D\varepsilon
  \]

- **Measured output**
- **Prediction Errors**
- **Known function of** \( \theta \)
Given the measurements equation:

\[ Y = F(\theta) + D_{\varepsilon} \]

It is possible to estimate \( \theta \) by means of the Prediction Error (PE) method:

\[ \hat{\theta}^{LS} = \arg \min_{\theta} V_N(\theta) \]

\[ V_N(\theta) = \frac{1}{N} D_{\varepsilon}^T D_{\varepsilon} = \frac{1}{N} [Y - F(\theta)]^T [Y - F(\theta)] \]

**Problem:** \( V_N(\theta) \) is in general non-convex.
Parametric approach

- If possible, physical laws are used to obtain the parametric representation of $f(w, \theta)$.

- When the physical laws are not well known or too complex, black-box parameterizations are used.

Fixed basis parameterization
Polynomial, trigonometric, etc.

Tunable basis parameterization
Neural networks
Fixed basis functions

\[ f(w, \theta) = \sum_{i=1}^{r} \alpha_i \sigma_i(w) \quad \theta = [\alpha_1 \cdots \alpha_r]^T \]

\( \sigma_i(w) \): Basis functions

**Problem:** Can \( \sigma_i \)'s be found such that:

\[ f(w, \theta) \underset{r \to \infty}{\to} f^o(w) \]
Fixed basis functions

For continuous \( f^o \), bounded \( W \subset \mathbb{R}^n \) and \( \sigma_i \) polynomial of degree \( i \) (Weierstrass):

\[
\limsup_{r \to \infty} \sup_{w \in W} \left| f^o(w) - f(w, \theta) \right| = 0
\]

Polynomial models
Fixed basis functions

\[ f(w, \theta) = \sum_{i=1}^{r} \alpha_i \sigma_i(w) \quad \theta = [\alpha_1 \cdots \alpha_r]^T \]

- **NARX models**: PE estimation of \( \theta \) is a linear problem:

\[
Y = L\theta + D\varepsilon
\]

\[
L = \begin{bmatrix}
\sigma_1(w^1) & \cdots & \sigma_r(w^1) \\
\vdots & \ddots & \vdots \\
\sigma_1(w^N) & \cdots & \sigma_r(w^N)
\end{bmatrix}
\quad Y = \begin{bmatrix}
y^2 \\
\vdots \\
y^{N+1}
\end{bmatrix}
\]

- **Least squares solution**: \( \hat{\theta}_{LS} = (L^T L)^{-1} L^T Y \)
Tunable basis functions

\[ f(w, \theta) = \sum_{i=1}^{r} \alpha_i \sigma(w, \beta_i) \]

\[ \theta = [\alpha_1 \ldots \alpha_r \beta_{11} \ldots \beta_{rq}]^T, \quad \beta_i \in \mathbb{R}^q \]

- One of the most common tunable parameterization is the one-hidden layer sigmoidal neural network.
Parametric models

- Model structure choice:
  - Basis functions
  - Number of Basis functions
  - Number of regressors

- Problem: curse of dimensionality
  The number of parameters needed to obtain “accurate” models may grow exponentially with the dimension $n$ of regressor space.

  More relevant in the case of fixed basis functions
Under suitable regularity conditions on the function to approximate, the number of parameters required to obtain “accurate” models grows linearly with $n$.

Estimation of $\theta$ requires to solve a non-convex minimization problem (even for NARX models).

Trapping in local minima
Nonlinear regression systems

Consider a nonlinear system in regression form:

\[ y^{t+1} = f(w^t) + d^{t+1} \]

where:

- \( w^t \): regressor. It defines the system structure:
  - \( w^t = [y^t \ y^{t-1} \ ... \ u^t \ u^{t-1} \ ...]^T \) \( \iff \) NARX
  - \( w^t = [f(w^{t-1}) \ f(w^{t-2}) \ ... \ u^t \ u^{t-1} \ ...]^T \) \( \iff \) NOE
  - \( w^t = [y^t \ y^{t-1} \ ... \ u^t \ u^{t-1} \ ... \ d^t \ d^{t-1} \ ...]^T \) \( \iff \) NARMAX

- \( u \): input signal.
- \( d \): noise acting on the system.
Nonlinear regression systems

The predictor of system $f$ is defined as:

$$\hat{y}_{t+1} = f(w^t)$$

where:

$$w^t = [y^t \ y^{t-1} \ ... \ u^t \ u^{t-1} \ ...]^T \iff \text{NARX}$$

$$w^t = [\hat{y}^t \ \hat{y}^{t-1} \ ... \ u^t \ u^{t-1} \ ...]^T \iff \text{NOE}$$

$$w^t = [y^t \ y^{t-1} \ ... \ u^t \ u^{t-1} \ ... \ \epsilon^t \ \epsilon^{t-1} \ ...]^T \iff \text{NARMAX}$$

$$\epsilon^t = y^t - \hat{y}^t \ : \ \text{prediction error}$$