Nonlinear system identification

Mario Milanese, Carlo Novara
Politecnico di Torino

Nonlinear system identification

- Consider a nonlinear system in regression form:

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

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

- The function f^o is unknown, but a finite set of noise-corrupted measurements of y^t and w^t is available:

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

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

Identification problem: find an estimate $\hat{f}$ of f^o

Nonlinear system identification

- Related problems :
  - *for a given estimate* $\hat{f} \cong f^o$

    evaluate the identification error $\left\| f^o - \hat{f} \right\|$
  - *find an estimate* $\hat{f} \cong f^o$

    minimizing the identification error

- The estimation error cannot be exactly evaluated since f^o is not known

- Need of prior assumptions on f^o and d^t for deriving finite a bound on this error

Nonlinear system identification

- Typical assumptions:
  - on system: $f^o \in \mathcal{F}(\theta) = \left\{ f(w,\theta) = \sum_{i=1}^{r} \alpha_i \sigma_i(w,\beta_i) \right\}$
  - on noise: iid stochastic noise

- Functional form of f^o required:
  - derived from physical laws
  - σ_i : basis function (polynomial, sigmoid,..)

- The parameters θ are estimated by means of the Prediction Error method using least squares

Parametric approach

$$f(w,\theta) = \sum_{i=1}^{r} \alpha_i \sigma\left(w, \beta_i\right)$$

- Given T noise-corrupted measurements of y^t and w^t,

$$\tilde{y}^2 = f(\tilde{w}^1, \theta) + d^1$$
$$\tilde{y}^3 = f(\tilde{w}^2, \theta) + d^2$$
$$\vdots$$
$$\tilde{y}^{T+1} = f(\tilde{w}^T, \theta) + d^T$$

$$Y = F(\theta) + D$$

Measured output

Known function

Error

Parametric approach

- **Prediction Error** estimate of θ:

$$\hat{\theta} = \arg\min_{\theta} V_T(\theta)$$ ← **least-squares**

$$V_T(\theta) = \frac{1}{T}\left\| Y - F(\theta) \right\|^2 = \frac{1}{T}\sum_{k=1}^{T}\left[\tilde{y}^{k+1} - f(\tilde{w}^k, \theta)\right]^2$$

- **Problem:** $V_T(\theta)$ 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

Polinomial, trigonometric, etc.

"Tunable" basis perametrization

Neural networks

"Fixed" basis functions

$$f(w,\theta)=\sum_{i=1}^{r}\alpha_i\sigma_i(w) \qquad \theta=\left[\alpha_1\cdots\alpha_r\right]'$$

$\sigma_i(w)$: Basis functions

- **Problem:** Can σ_i's be found such that

$$f(w,\theta)\xrightarrow[r\to\infty]{} f^o(w) \qquad ?$$

"Fixed" basis functions

- For continuous f^o, bounded $W \subset \Re^n$ and σ_i polynomial of degree i (Weierstrass):

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

Polynomial NARX models

"Fixed" basis functions

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

- Estimation of θ is a convex problem: $Y = L\theta + D$

$$L = \begin{bmatrix} \sigma_1(\tilde{w}_1) & \cdots & \sigma_r(\tilde{w}_1) \\ \vdots & \ddots & \vdots \\ \sigma_1(\tilde{w}_N) & \cdots & \sigma_r(\tilde{w}_N) \end{bmatrix} \qquad Y = \begin{bmatrix} \tilde{y}^2 \\ \vdots \\ \tilde{y}^{N+1} \end{bmatrix}$$

- Least-squares solution:

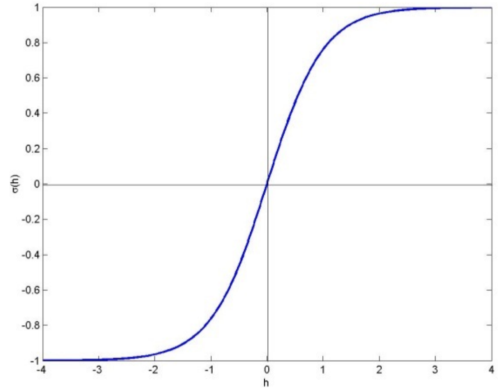
$$\boxed{\hat{\theta} = (L'L)^{-1} L'Y}$$

"Tunable" basis functions

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

$$\theta = [\alpha_1 \cdots \alpha_r \;\; \beta_{11} \cdots \beta_{rq}]', \quad \beta_i \in \Re^q$$

- One of the most common "tunable" parameterization is the one-hidden layer sigmoidal neural network



"Tunable" basis functions

- The parameters β_i give more flexibility to the model, possibly providing a more accurate estimate

- On the other hand, parameter estimation require to solve a non-convex optimization problem, due to the fact that the parameters appear nonlinearly:

$$Y = F(\theta) + D$$

nonlinear in θ

Parametric models

- Model structure choice:
  - type of basis functions
  - Number r of "Basis" functions
  - Number n of regressors

The complexity of these problems may be exponential in n.

- **Problem: curse of dimensionality**

  The number of parameters r needed to obtain "accurate" models may grow exponentially with the dimension n of regressor space

↑

More relevant in the case of "fixed" basis functions

"Tunable" basis functions

- Under suitable regularity conditions on the function to approximate, the number of parameters r required to obtain "accurate" models grows linearly with n

- The estimation of θ requires to solve a non-convex minimization problem

↓

Trapping in local minima

One-step/multi-step prediction, simulation

- Notation: $\hat{y}^t$ = predicted output; $\tilde{y}^t$ = measured output; $\tilde{u}^t$ = measured/known input.
- One-step prediction:

$$\hat{y}^{t+1} = f(\tilde{y}^t, \tilde{y}^{t-1}, \ldots, \tilde{u}^t, \tilde{u}^{t-1}, \ldots).$$

- Multi-step prediction and simulation:

$$\hat{y}^{t+1} = f(\hat{y}^t, \hat{y}^{t-1}, \ldots, \tilde{u}^t, \tilde{u}^{t-1}, \ldots).$$

- Model identification consists in minimizing
  - the prediction error → Nonlinear AutoRegressive with eXternal input (NARX) models; linear case: ARX models;
  - the simulation error → Nonlinear Output Error (NOE) models; linear case: OE models.
- (N)OE models may be more effective than (N)ARX models for simulation or multi-step prediction tasks.