Set Membership identification of nonlinear models

Carlo Novara

Dip. Elettronica e Telecomunicazioni
Politecnico di Torino

Outline

Identification problem

- Consider a nonlinear system described in NARX form:

$$y_{k+1} = f^{o}(w_k) + d_k$$
$$w_k \doteq (y_k, \ldots, y_{k-n_a+1}, u_k, \ldots, u_{k-n_b+1})$$

  $u_k \in \mathbb{R}^{n_u}$: input
  $y_k \in \mathbb{R}^{n_y}$: output
  $w_k \in \mathbb{R}^{n}$: regressor, $n = n_a n_y + n_b n_u$
  $d_k \in \mathbb{R}^{n_y}$: noise accounting for input and output disturbances/errors
  $k = 0, 1, 2, \ldots$: time index.

- The NARX model structure is quite general: A large number of real-world dynamic systems can be captured by this structure.

Identification problem

- Consider a nonlinear system described in NARX form:

$$y_{k+1} = f^o\left(w_k\right) + d_k$$

$u_k \in \mathbb{R}^{n_u}$: input
$y_k \in \mathbb{R}^{n_y}$: output
$w_k \in \mathbb{R}^{n}$: regressor
$d_k \in \mathbb{R}^{n_y}$: disturbance/noise.

- Suppose that:
  - d_k is unknown
  - f^o is unknown
  - a set of data $\mathcal{D} \doteq \{\tilde{y}_{k+1}, \tilde{w}_k\}_{k=1}^{N}$ is available.

Identification problem. Find an accurate (in some sense) estimate $\hat{f}$ of f^o.

Identification problem

Related important problems:

- ◇ For an estimate $\hat{f} \cong f^o$, evaluate the *identification error* $\left\| f^o - \hat{f} \right\|$.
- ◇ Find an estimate that **minimizes the identification error**.

- However, the identification error cannot be evaluated, since f^o is unknown.
- Need of **prior assumptions** on the system (represented by f^o) and on the noise d_k to derive a finite bound on this error.

Parametric statistical approach

Classical assumptions:

- **Noise:** Statistical assumptions, like zero-mean, i.i.d., Gaussian, ...
- **System:** f^o belongs to a set of parametrized functions:

$$f^o \in \mathcal{F}_P(\theta) \doteq \left\{ f(w,\theta) = \sum_{i=1}^{m} \alpha_i \sigma_i(w,\beta_i) \right\}$$

$$\theta = (\alpha_1, \dots \alpha_r, \beta_1, \dots \beta_r). \quad \square$$

- Choice of the basis functions σ_i:
  - based on physical laws (when available);
  - "universal" approximators (polynomial, trigonometric, sigmoidal, ...).
- The parameters in θ are typically estimated by solving an optimization problem (e.g., minimization of the prediction error).
  - Parameter estimation is often called **learning**.

Nonlinear Set Membership approach

- In several real-world situations:
  - The statistical assumptions may not hold or not be reliable.
  - The physical laws may be not well known or too complex.
- The Nonlinear Set Membership (NSM) approach is based on somewhat weaker assumptions:

NSM assumptions:

◇ **Noise:** bounded as $\|d_k\| \leq \epsilon, \ \forall k$.

◇ **System:** f^o belongs to a set of functions with gradient (or Jacobian) bounded by a constant γ:

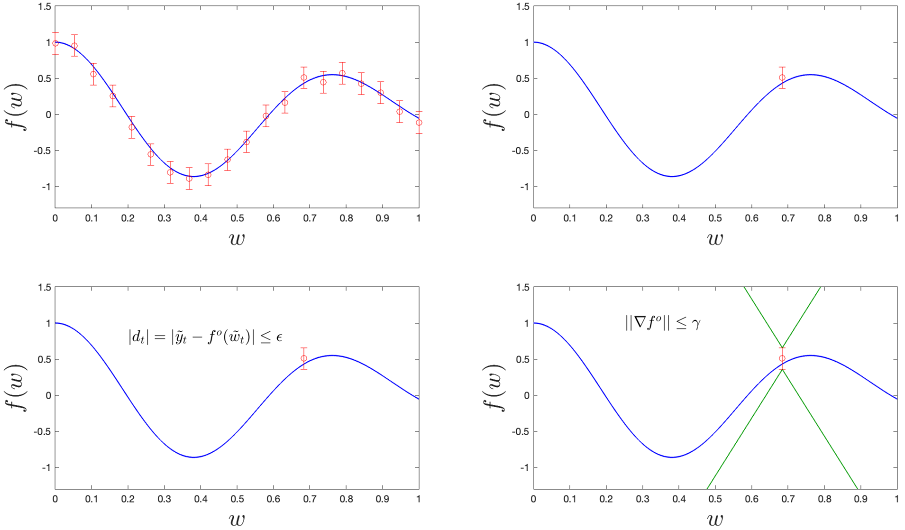
$$f^o \in \mathcal{F}_S(\gamma) \doteq \left\{ f \in C^1 : \|\nabla f(w)\| \leq \gamma, \ \forall w \in W \right\}$$

$$W = \text{function domain} = \text{bounded set of } \mathbb{R}^n. \quad \square$$

- The generalization from C^1 to Lipschitz is straightforward.
- γ and ϵ are estimated from data by means of a validation criterion.

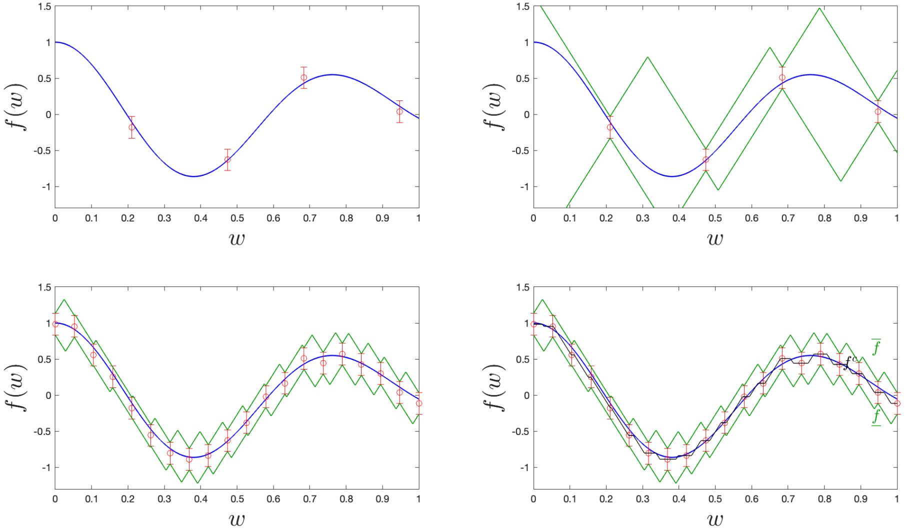
Nonlinear Set Membership approach

Basic idea ($w \in \mathbb{R}$) - information utilization



Nonlinear Set Membership approach

Basic idea ($w \in \mathbb{R}$) - uncertainty bounds and central estimate



Nonlinear Set Membership approach

Uncertainty bounds and central estimate

- The bounds and the central estimate can be computed in closed-form in the general case $w \in \mathbb{R}^n$:

$$\overline{f}(w) \doteq \min_{k=1,\ldots,N} \left(\tilde{y}_{k+1} + \epsilon + \gamma \left\| w - \tilde{w}_k \right\|\right)$$

$$\underline{f}(w) \doteq \max_{k=1,\ldots,N} \left(\tilde{y}_{k+1} - \epsilon - \gamma \left\| w - \tilde{w}_k \right\|\right)$$

$$f^c(w) = \frac{1}{2}\left(\underline{f}(w) + \overline{f}(w)\right).$$

- In the following, we will see that:
  - $\underline{f}$ and $\overline{f}$ are *optimal uncertainty bounds*: they are the tightest bounds on the unknown function f^o that can be obtained from the available information.
  - f^c is an *optimal estimate*: it minimizes the worst-case identification error (to be defined later).

Nonlinear Set Membership theory

Feasible System Set

- All the information (prior and data) available at time N is summarized by the *Feasible System Set*.

Definition. Feasible System Set:

$$FSS \doteq \{f \in \mathcal{F}_S(\gamma) : |\tilde{y}_{k+1} - f(\tilde{w}_k)| \leq \epsilon,\ k = 1, \ldots, N\}. \quad \square$$

- FSS is the set of all systems compatible with the prior information (ϵ and γ) and data.
- In other words, FSS is the set of all systems $f \in \mathcal{F}_S(\gamma)$ that could have generated the data.

Nonlinear Set Membership theory

Validation of prior assumptions

- If $FSS = \emptyset$ it means that no system exists compatible with prior assumptions and data $\Rightarrow$ the assumptions are falsified by data.
- If $FSS \neq \emptyset$ at least one system exists compatible with prior assumptions and data $\Rightarrow$ the assumptions are validated by data.

Definition. The assumptions are considered validated if $FSS \neq \emptyset$. $\square$

- The fact that the priors are validated by the present data does not exclude that they may be invalidated by future data (Popper, Conjectures and Refutations: the Growth of Scientific Knowledge", 1969).

Theorem. Conditions for $FSS \neq \emptyset$ are

◇ Necessary: $\overline{f}(\tilde{w}_k) \geq \underline{f}(\tilde{w}_k)$ $k = 1, \ldots, N$.

◇ Sufficient: $\overline{f}(\tilde{w}_k) > \underline{f}(\tilde{w}_k)$ $k = 1, \ldots, N$. $\square$

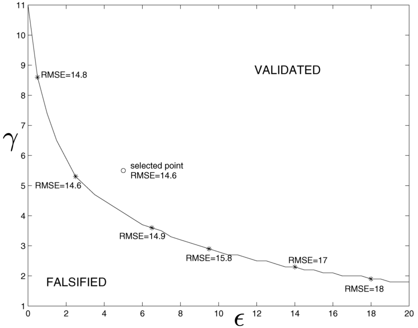
Nonlinear Set Membership theory

Validation of prior assumptions

- Using the above theorem, the following curve can be constructed:

$$\gamma_{min}(\epsilon) \doteq \inf_{FSS \neq \emptyset} \gamma.$$

- For each ϵ, it gives the minimum γ ensuring assumption validation.
- It separates the validated and falsified parameter regions.
- The curve, together with some accuracy criterion, can be used to choose the values of ϵ and γ.



Nonlinear Set Membership theory

Optimality approximation

- Let $\hat{f}$ be an approximation of the unknown "true" function f^o.

Definition. Worst-Case (WC) identification error of $\hat{f}$:

$$E(\hat{f}) \doteq \sup_{f \in FSS} \left\| f - \hat{f} \right\|. \quad \square$$

- The error is measured using a L_p functional norm, given by

$$\|f\|_p \doteq \begin{cases} \left[\int_W \|f(w)\|_p^p \, dw \right]^{\frac{1}{p}}, & p < \infty, \\ \operatorname{ess\,sup}_{w \in W} \|f(w)\|_\infty, & p = \infty. \end{cases}$$

Definition. An approximation f^* is *optimal* if $E(f^*) = \inf_f E(f) = r_I$.
r_I = *radius of information*; it is the minimum WC error achievable. $\square$

Nonlinear Set Membership theory

Optimal model

Assumption: $FSS \neq \emptyset$.

Theorem.

i) f^c is an *optimal estimate* for any L_p norm.

ii) The radius of information is given by $r_I = \frac{1}{2} \left\| \overline{f} - \underline{f} \right\|_p$. $\square$

Theorem. $\underline{f}$ and $\overline{f}$ are *optimal bounds*: they are the tightest bounds on f^o that can be derived on the basis of the available information:

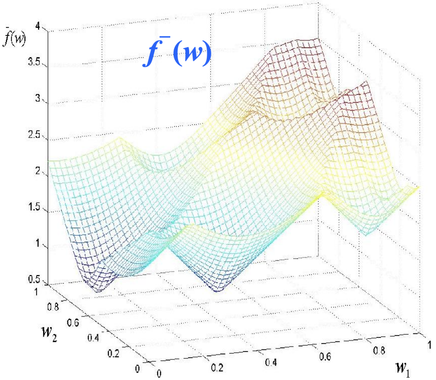
$$\overline{f}(w) = \sup_{f \in FSS} f(w), \; \forall w \in W$$

$$\underline{f}(w) = \inf_{f \in FSS} f(w), \; \forall w \in W. \quad \square$$

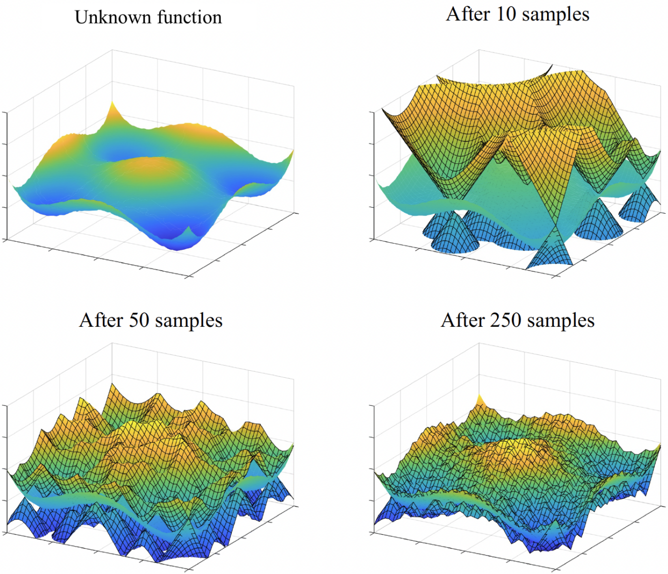
Mathematical properties - uncertainty bounds

Properties of $\underline{f}$ and $\overline{f}$:

- Piecewise conic functions.
- Lipschitz continuous with constant γ (they are not C^1).
- Differentiable almost everywhere.

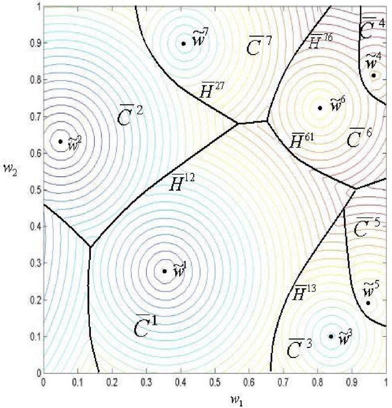


Mathematical properties - uncertainty bounds

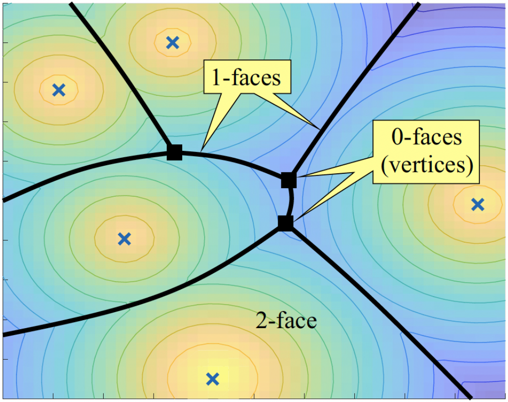


Mathematical properties - Hyperbolic Voronoi Diagrams

- The projections of the cones intersections generate the so-called Hyperbolic Voronoi Diagrams (HVD).
- HVDs define a partition of the function domain, featuring faces of different dimensions (form 0 to n).
- They are generalizations of standard Voronoi Diagrams (Edelsbrunner, Combinatorial Geometry, Springer, 1987).



Mathematical properties - Hyperbolic Voronoi Diagrams

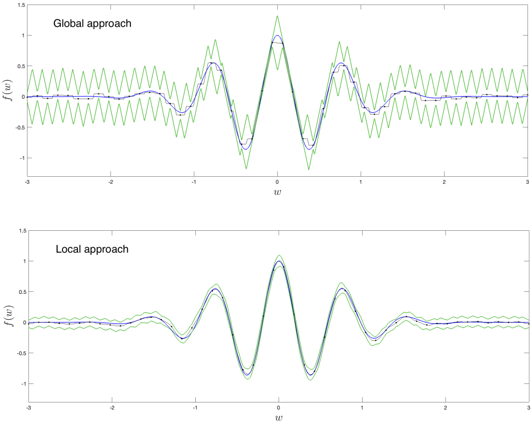


NSM local approach

- A global bound on the norm of the unknown function gradient as been assumed so far: $\|\nabla f^o(w)\| \leq \gamma$, $\forall w \in W$.
- **Problems:**
  - $\nabla f^o(w)$ may drastically change in function of w. A global gradient bound does allow an effective adaptation.
  - In the case of low number of identification data, the uncertainty region defined by $\underline{f}$ and $\overline{f}$ may be quite large.
- Simple method to overcome such problems:
  - Identify a function f^a approximating f^o, using any approach (e.g., polynomials, neural networks, support vector machines, ...).
  - Apply the NSM approach to the *residue function* $f_\Delta \doteq f^o - f^a$, using the data $\mathcal{D}_\Delta \doteq \{\tilde{y}_{k+1} - f^a(\tilde{w}_k), \tilde{w}_k\}_{k=1}^N$.
- The global bound $\|\nabla f_\Delta\| \leq \gamma_\Delta$ implies local bounds on $\|\nabla f^o\|$:

$$\|\nabla f_a(w)\| - \gamma_\Delta \leq \|\nabla f_o(w)\| \leq \|\nabla f_a(w)\| + \gamma_\Delta.$$

NSM local approach



Remark: Thanks to the local approach, NSM identification can be combined with any other modeling method, allowing an easy evaluation of tight uncertainty bounds.